VOL. 11 No. 9, pp. 253-280 MARCH 1950

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

A PRESERVATION RECTIFIER WITH ELECTRONICALLY STABILIZED CHARGING VOLTAGE

by E. CASSEE.

621.314.6.076.7:621.316.261

In almost every field of electrical technology there are nowadays installations in which electronic control is applied. The present article explains how with such a control a rectifier can be given a characteristic highly suitable for preserving or maintaining the charge of an accumulator battery.

Preservation charge

In the course of time electricity supplies have been so improved by developments as to form a source of energy quite reliable for industrial and domestic purposes notwithstanding the breakdowns that are still apt to occur at times. Nevertheless there are a number of cases where even a very short interruption of the current supply may have unpleasant and even fatal consequences; we have only to think of telephonic, telegraphic and telex communications and the lighting of theatres, traffic tunnels, operating theatres, etc. Wherever this is the case an emergency supply system, consisting for instance of a battery of accumulators, is therefore installed.

An article which appeared earlier in this journal [1]) dealt at length with what was termed the "hygienics" of the battery, i.e. its maintenance. There it was shown that a single battery kept under what is called a preservation charge has great advantages compared with the former system generally employed for telephone and similar plants where two batteries are charged and discharged alternately. As the first of these advantages is to be mentioned the much longer life of the battery, and further the saving in initial cost, space and upkeep when instead of two batteries only one is used. Other advantages lie in the higher efficiency of preservation charging and the absence of any formation of gases and noxious vapours.

Preservation charging means that the battery is kept fully charged by feeding it continuously with a low current just sufficient to compensate the internal losses. Thus the battery is permanently connected in parallel with the source of the charging current (and, for instance in the case of a telephone exchange, also with the direct-current-consuming network, in which case the battery is said to be kept in a "floating" condition). This charging-current source, which, as an example, may be a rectifier fed from the A.C. mains, has to answer very special requirements. It has to feed the battery continuously with a small current just sufficient to make up for the internal losses of the battery, this being achieved by maintaining a voltage of 2.1 to 2.2 V per cell [2]). In the case of a telephone exchange the charging current has also to supply the current required for the working of the plant (this current is not usually supplied by the battery, except in a special case to which we shall revert later, and of course during a failure of the A.C. mains).

The manner in which a rectifier possessing these properties — "a preservation rectifier" — can be devised has already been mentioned in the article referred to in footnote [1]) and also in an article of an earlier date [3]). In the simplest case the principle is as represented in *fig. 1*: a set of

[1]) H. A. W. Klinkhamer, Emergency supply systems with accumulator batteries, Philips Techn. Rev. **9**, 231-238, 1947.

[2]) Here we are speaking of lead accumulators. Our arguments apply, mutatis mutandis, in broad lines also for nickel-iron cells.

[3]) H. A. W. Klinkhamer, A rectifier for small telephone exchanges, Philips Techn. Rev. **6**, 39-45, 1941.

valves (for instance connected for single-phase full-wave rectification) is fed from the series-connected secondary windings of two transformers, one of which (T_2) is connected direct to the A.C. mains while

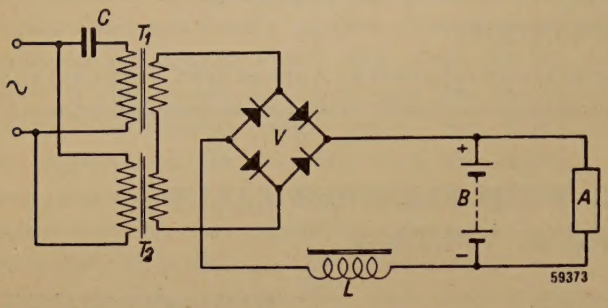


Fig. 1. Circuit of a preservation rectifier previously described (see footnotes [1]) and [3])). T_1 = transformer with strongly saturated core, T_2 = ordinary transformer, C = capacitor in series with the primary of T_1, V = single phase, full wave rectifying circuit, L = smoothing choke, B = battery, A = plant.

the other (T_1) is connected to the mains via a capacitor. It is essential that the iron core of the latter transformer should be strongly saturated. For the working of this system reference may be made to the articles [1]) and [3]) already quoted. In the latter article it was stated that, with a certain form of application of the system, in a wide range of the currents supplied by the rectifier (0.3 to 3 A) the charging voltage (i.e. the battery voltage) varies by only a few per cent (from 63 to 60 V). Furthermore the charging voltage is very little affected by variations in the mains voltage, a fluctuation of 10% causing a change in the charging voltage of at most 3%.

In three respects, however, the rectifiers working according to this system have less favourable properties. In the first place there is the fact — already mentioned in the article quoted in footnote [3]) — that the charging voltage is to a rather considerable extent dependent upon the frequency of the A.C. mains, a deviation of 1% from the rated frequency causing a change of about 1.6% in the charging voltage. Now since the war frequency variations up to about 4% often occur during a number of hours per day, at least in the networks in Western and Central Europe. The resultant variations in the charging voltage, together with those due to mains voltage fluctuations and to fluctuations in the consumption of the direct current, may be greater than is permissible for the proper preservation of the battery.

The second drawback is that the charging voltage is dependent upon the voltage loss in the valves, which is not constant in the course of time. In selenium rectifiers, for instance, the voltage loss increases with the current passed through, whilst, moreover, at a certain current intensity it becomes gradually greater, especially during the first period after the valve has been taken into use. It is of course possible to compensate changes in the voltage loss occurring with a given strength of current by connecting the valve to a different transformer tapping, but this is only a makeshift solution. In gas-filled rectifying valves the voltage loss (arc voltage) varies only slightly with the current but it nevertheless sometimes shows irregular fluctuations which cannot be compensated at all.

The third drawback of rectifiers according to the principle of fig. 1 is not of a fundamental nature but concerns the manufacture. The characteristic of these rectifiers representing the charging voltage as a function of the direct current is highly dependent upon the shape of the *B-H* curve of the iron in the transformer T_1, particularly in the region of strong saturation. Now the prescriptions generally applied for the testing of transformer sheets say nothing about this and transformer sheets from different batches show considerable differences, which of course give rise to difficulties in manufacture.

The increasing demand for preservation rectifiers has therefore made it necessary to find a system that is free of these drawbacks.

Preservation rectifier with electronic control

Stabilization of the charging voltage

The solution has been found in a system where the battery voltage itself performs a controlling action by regulating the D.C. magnetization of a transductor. A transductor is a choke whose impedance for alternating current can be varied by means of an auxiliary direct current (the control current) flowing through a separate winding on the core of the A.C. coil. With a transductor it is possible to arrange for a small D.C. power to control a much greater A.C. power or, if the alternating current is rectified, a much greater D.C. power.

The circuit is shown in *fig. 2*, where *Td* represents the transductor. When, for example, the D.C. output I_0 of the rectifier increases then the charging voltage tends to drop. However, by making the control current I_a of the transductor dependent upon the charging voltage — by means of the circuit denoted by F — the drop of this voltage can be made to increase the control current to such an extent that the impedance of the A.C. coil — and thus the voltage drop in that coil — decreases far enough to compensate practically the drop of the charging voltage.

For this purpose the control current cannot be taken from the battery direct, in the first place because the permissible fluctuations in the battery voltage are so small that they would have very

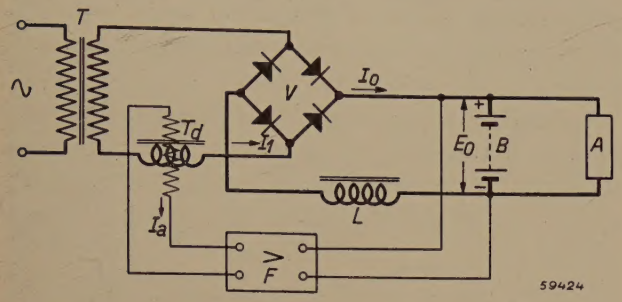


Fig. 2. Circuit of a preservation rectifier with a transductor (*Td*), the control current of which, I_a, is supplied by an amplifier *F* controlled by the charging voltage E_0. *T* = supply transformer. *A*, *B*, *L* and *V* as in fig. 1.

little effect in the transductor, and in the second place because this effect would be just in the wrong direction: for instance, when the battery voltage increases the control current would likewise increase, and, as already shown, this would result in a further increase of the battery voltage. That is why the battery voltage fluctations have to be amplified (in *F*, fig. 2) before they control the transductor, this amplification being accompanied by the necessary inversion of sign.

This amplification can be brought about in various ways. The simplest method is with electronic valves, for instance in the manner indicated in *fig. 3*, where the charging voltage E_0 is compared

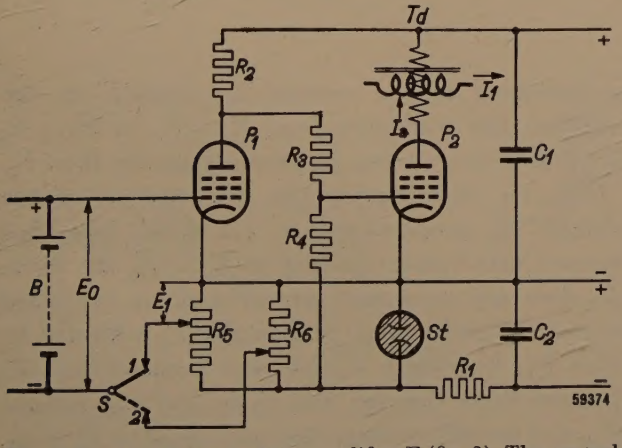


Fig. 3. Circuit of a part of the amplifier *F* (fig. 2). The control-grid voltage of the pentode P_1 (EF 41) is the difference between the battery voltage E_0 and a constant reference voltage E_1 adjustable with the potentiometer R_5. R_1 = current-limiting resistor of the voltage reference tube *St* (85 A1) which keeps E_1 constant. R_2, R_3 and R_4 are resistors for the intervalve coupling between P_1 and P_2 (EL 60), the anode current of which, I_a, controls the transductor *Td* (fig. 2). C_1, C_2 = smoothing capacitors of two small auxiliary rectifiers (not drawn). With the switch *S* one can change over to a higher reference voltage (for rapid charging) adjustable with a potentiometer R_6.

with a reference voltage E_1 that can be varied with the aid of a potentiometer R_5 and kept constant by means of a voltage reference tube *St*, fed from a small auxiliary rectifier. Variations in the potential difference $E_0 - E_1$ are amplified by the valves P_1 and P_2; in this case the anode current of P_2 is at the same time the control current I_a of the transductor. It can easily be seen that a change in the charging voltage changes the control current in the right sense.

With this system any change of the charging voltage is counteracted, no matter whether it arises from a variation of the voltage loss in the valves, from fluctuations of the mains voltage or from mains frequency variations. In how far the charging voltage can be kept constant in a practical case may be judged from *fig. 4*, representing the charging voltage as a function of the D.C. output current at different mains voltages. These characteristics have been plotted from measurements taken with a preservation rectifier type 3067 (*fig. 5*) for 64.5 V (30 lead cells), maximum 15 A, designed according to the system of fig. 3. With a frequency deviation of 4% or a mains voltage variation of 10% the change in the charging voltage is less than 0.5%. Thus not only the dependency upon the frequency (the main point at issue) but also the dependency upon the mains voltage is appreciably reduced in comparison with the former system. With the frequency and mains voltage fluctuations just given and with currents between 10% and 100% of the nominal rating, the total changes of the charging voltage remain between +1 V and −1 V.

In the article quoted in footnote [1]) it was described how very little dependency upon mains voltage fluctuations can be obtained also with preservation rectifiers according to the system of fig. 1. There use was made of the remarkable feature of these rectifiers that the charging voltage varies in the opposite sense to the mains voltage. By connecting such a rectifier in series with an ordinary rectifier, the two voltages of which act in the same direction, compensation can be brought about to a high degree.

As to the bends seen in the characteristics at small and at large currents, these will be referred to again farther on.

Regarding the question of the transformer sheet, with the transductor no difficulties arise of the nature of those encountered with the transformer T_1 (fig. 1). This is due in part to the manner in which the new rectifier works in bringing about the smallest possible difference between the charging voltage and the given, constant, reference voltage.

It is no exaggeration to say that the functioning of this rectifier stands or falls with the constancy of the reference voltage, thus with the properties of the voltage reference tube (*St*, fig. 3). A tube has therefore been used (type 85 A1) which shows only very small deviations from the rated voltage (85 V) either during its lifetime or with varying currents or temperature fluctuations; these deviations are never more than 1 V and mostly not even more than a few tenths of a volt. This tube, which works with a glow discharge in neon (average current 4 mA), has a carefully prepared cathode of molybdenum. A layer of the same metal is precipitated by atomization on the inner wall of the glass bulb and prevents the release of impurities from the glass. For further particulars reference may be made to another article published in this Journal [4]).

The potentiometer R_5 (fig. 3) allows of some regulation of the charging voltage, for instance between 2.10 V and 2.30 V per cell in the flat part of the characteristic.

Current limiting

The sharp bend of the characteristic at a high value of the current (fig. 4) is of great importance, for it means that the current I_0 supplied by the rectifier is limited. When the working of a plant requires from time to time a current higher than this threshold value then the difference has to be made up by the battery; thus the rectifier itself need not be calculated for these higher currents. It has to be explained, however, how this limitation of current comes about — it occurs also with preservation rectifiers of the old type, but in a different way — since the arrangement according to fig. 3 is not sufficient.

This effect is obtained by arranging for the grid voltage of the valve P_1 to consist, from a certain strength of current onwards, not of $E_0 - E_1$ (fig. 3) but of $E_2 - E_1$, where E_2 is a direct voltage

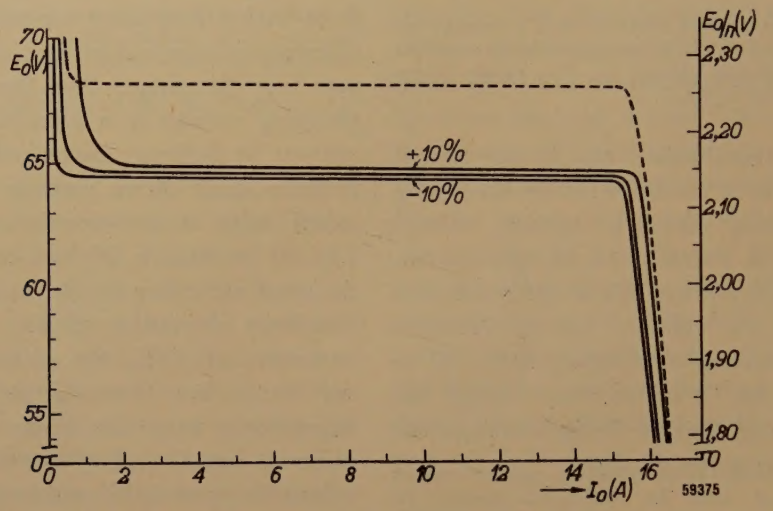


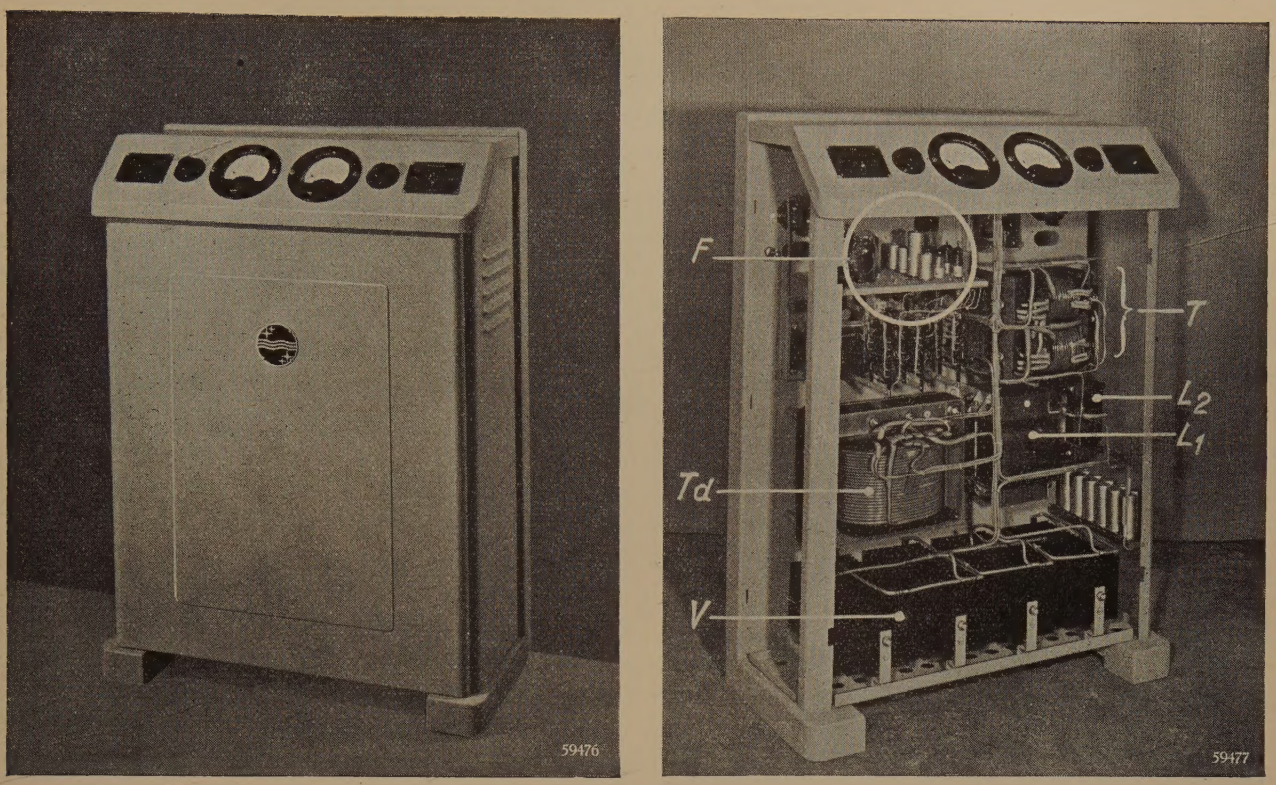
Fig. 4. Charging voltage E_0 as a function of the direct current I_0 of the preservation rectifier type 3067 for 64.5 V (30 lead cells), max. 15 A. Fully drawn lines: preservation charging currents respectively with 10% too low, with normal and with 10% too high mains voltage. Dotted line: rapid charging. On the right-hand scale the voltage per cell E_0/n (n = number of cells in series).

proportional to the current I_0 supplied by the rectifier; the transition from $E_0 - E_1$ to $E_2 - E_1$ is brought about when E_2 becomes greater than E_0.

Fig. 6a (disregarding the derivation of the voltage E_2 proportional to I_0) shows how this process takes place: so long as $E_2 < E_0$ the diode D_1 does not pass any current and, as explained above, the grid voltage of the valve P_1 is formed by $E_0 - E_1$, but when E_2 becomes greater than E_0 current flows through the diode and the resistor R_7, and the grid voltage becomes equal to $E_2 - E_1$. Thus as I_0 increases this grid voltage now changes in the positive direction and this — as may easily be seen from fig. 3 — reduces the control current. As a consequence the voltage drop in the A.C. coil of the transductor is greatly increased and this results in a rapid drop of the charging voltage.

It now remains to be seen how the voltage E_2 proportional to I_0 is obtained. The alternating

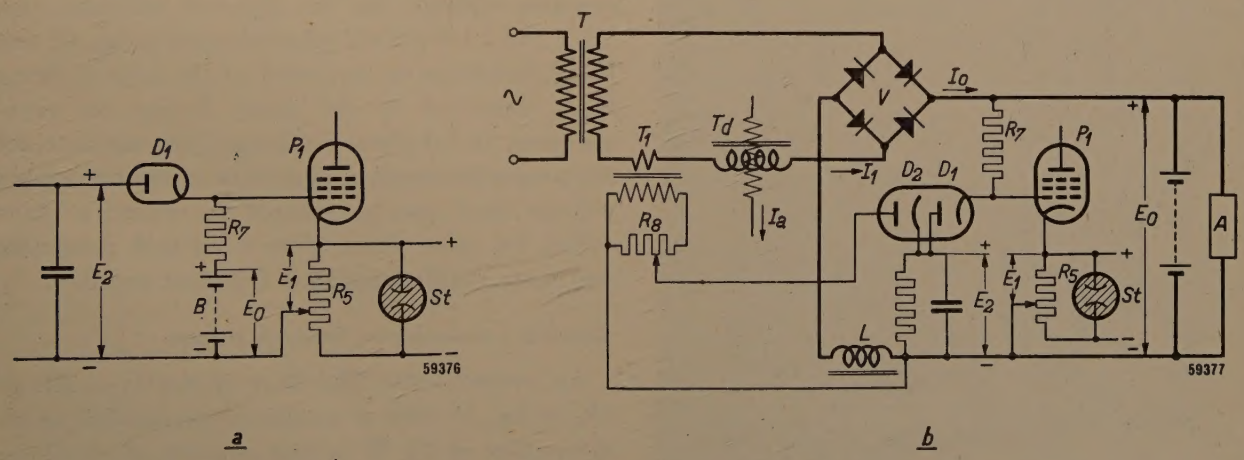
[4]) T. Jurriaanse, A voltage-stabilizing tube for very constant voltage, Philips Techn. Rev. 8, 272-277, 1946.



a *b*

Fig. 5. Electronically controlled preservation rectifier type 3067 for 64.5 V, max. 15 A. *a*) closed, *b*) open. L_1, L_2, two smoothing chokes. Other lettering the same as in fig. 2.

current I_1 taken up by the rectifier is proportional to I_0 and flows through the primary coil of a small auxiliary transformer (T_1, fig. 6*b*), inducing in the secondary winding a voltage which is proportional to I_1 and thus also proportional to I_0. Rectification of this voltage, with the aid of a diode D_2, and smoothing yields the desired D.C. voltage E_2. By making this voltage adjustable, for instance with the potentiometer R_8, one can adjust, if required, the value to which I_0 is limited.



a *b*

Fig. 6. Circuit for limiting the current. *a*) When the voltage E_2 exceeds E_0 current flows through the diode D_1 and the resistor R_7, and the control-grid voltage of P_1 (fig. 3) is then no longer $E_0 - E_1$ but $E_2 - E_1$. The other letters are the same as in fig. 3. *b*) The voltage E_2 is obtained by rectification (diode D_2) and smoothing of the alternating voltage (adjustable with a potentiometer R_8) of the secondary of the transformer T_1. Flowing through the primary of T_1 is an alternating current I_1 proportional to the direct current I_0. Thus E_2 is also proportional to I_0. The diodes D_1 and D_2 are contained in one bulb (EB 41). The other letters are the same as in the foregoing diagrams.

The working point of the rectifier is shifted to the bend in the characteristic (fig. 4) at low current — this bend also occurs with preservation rectifiers of the old type — when the battery voltage rises (for instance, due to there being no load for a long time) so far that the valve P_2 is cut off and thus the control current drops to about zero. The transductor then acts as an ordinary choke (with constant, high impedance) and the rectifier as an ordinary rectifier. The current at which this bend in the characteristic occurs can be made smaller by using a transductor designed for a greater self-inductance (without control current), but this would necessarily be heavier and more expensive. With the normal design of transductor this bend occurs at a current which is so small that usually the self-discharge of the battery, together with the basic consumption nearly always present (relay coils, etc), is already sufficient to prevent the rectifier from coming into action in the steep part of the characteristic. Should this load be insufficient — in which case undesired "gasing" of the battery might arise — then a small, additional, basic load can be applied, for instance in the form of a resistor shunted across the battery and the rectifier, or else a resistor or a choke shunted across the A.C. terminals of the rectifying circuit.

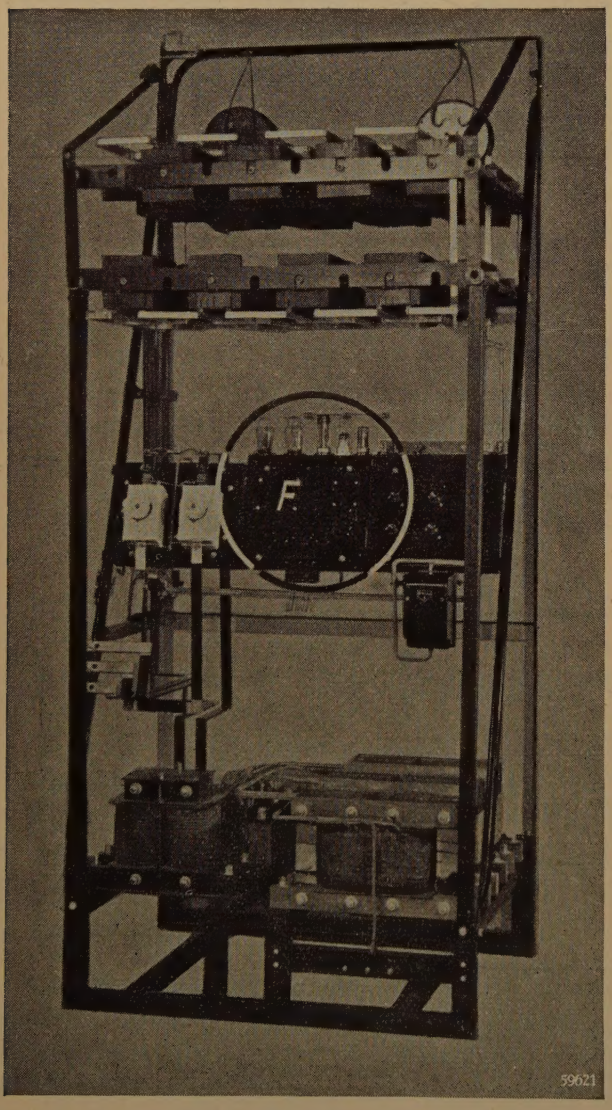


Fig. 7. Back view of a preservation rectifier for 64.5 V, max. 105 A, destined for large telegraph stations. The rectifier is of the three-phase construction but otherwise it functions in the same way as type 3067 (fig. 5). The amplifier is again denoted by *F*.

Charging of the battery after interruption of the mains

When the battery has performed its duty as a reserve source of energy during interruption of the mains supply then upon the latter being restored it has to be rapidly recharged so as to be ready for another possible emergency. At the beginning of this charging the cells have a voltage of 2.0 V or less and the charging current is therefore of the maximum value (e.g. 15 A in the case of the rectifier 3067, the characteristics of which are represented in fig. 4). With a normal characteristic, the horizontal part of which lies at, say, 2.15 V per cell, the charging current would drop to a small value (about 1 A in the case of the type 3067) already when that voltage is reached and the charging would therefore take a long time, so that the full capacity of the battery would not become available for a considerable time. To reduce this charging time provision has been made for rapid charging, simply by increasing temporarily the reference voltage E_1, thus raising the charging voltage to a higher level corresponding, for instance, to 2.3 V per cell. The charging current is then maintained at its maximum value until this higher voltage is reached and the battery is therefore charged more rapidly. The change-over to rapid charging is brought about by turning a switch (S, fig. 3) to position *2*, thereby replacing the potentiometer R_5 by the potentiometer R_6, by means of which the charging voltage can be adjusted between, say, 2.2 V and 2.4 V per cell (dotted curve in fig. 4), this being desired so as to permit of the battery being kept connected to the plant during the rapid charging. In telephone exchanges, for instance, for the proper functioning of certain relays the supply voltage must not be allowed to exceed a given value; for other plants there is no such restriction and more rapid charging is therefore possible.

Auxiliary transductor for high powers

An output valve like that of the type EL 60 (P_2 in fig. 3) with a maximum permissible anode dissipation of 25 W is just capable of supplying the control current for the transductor of a preservation rectifier for 64.5 V, 15 A. For rectifiers of a higher power a larger output valve could be chosen, but for rectifiers having a power of, say, 64.5 V, 105 A (*fig. 7*) we have followed another way, namely a sort of cascade circuit of two transductors. Here

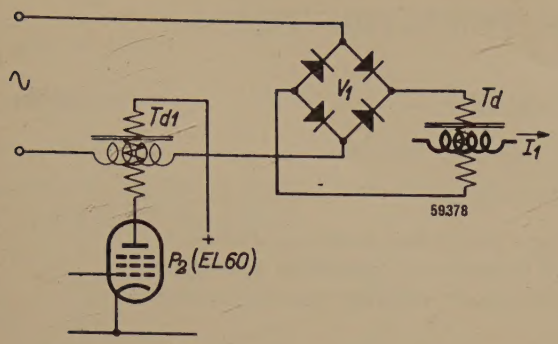


Fig. 8. Cascade circuit of an auxiliary transductor Td_1 and the main transductor Td. V_1 = auxiliary rectifier. In this manner a main transductor of high power can be controlled with a relatively small output valve (EL 60).

the valve EL 60 controls a small auxiliary transductor (Td_1, *fig. 8*). This again controls an auxiliary rectifier (V_1), which in turn supplies the control current for the main transductor (Td). For very high powers the cascade circuit can be extended, if necessary, by more auxiliary transductors.

Summary. The fact is recalled that for the power supply of telephone exchanges, for instance, a single battery with preservation charge is better than two batteries alternately charged and discharged, as used to be the case. Preservation charging means that the battery is permanently connected to a charger and that its voltage is maintained between 2.1 and 2.2 V per cell (for lead cells). As source of the charging current use can be made of a rectifier with a constant charging voltage. Here a system is described for such a preservation rectifier where the charging voltage is compared with a voltage which is to be regarded as being constant and which is stabilized by a voltage reference tube with narrow tolerance. The difference of the two voltages is amplified by electronic valves and controls the D.C. magnetization of a transductor (a choke the impedance of which is variable by means of D.C. saturation) connected in series with the A.C. side of the rectifier. Between wide limits of the D.C. output the charging voltage is very little affected by the voltage drop in the valves (thus also by the direct current), by the mains voltage or by the frequency. In this and other respects there is a great improvement compared with preservation rectifiers of an older system. An addition to the controlling mechanism makes it possible to limit the direct current to a certain adjustable value. By means of a switch a higher charging voltage can be chosen for rapidly recharging a battery after it has been wholly or partly discharged in the event of a prolonged interruption of the mains supply. In the controlling mechanism a pentode EL 60 (with a maximum permissible anode dissipation of 25 W) is used as output valve; this supplies direct the control current for the transductor of a rectifier for, say, 64.5 V, 15 A. In the case of higher powers, instead of choosing a larger output valve it is often more advantageous to arrange for the output valve to control an auxiliary transductor which in turn controls the main transductor (if necessary applying more than one auxiliary transductor).

THE THEORY OF SAMPLING INSPECTION PLANS

by H. C. HAMAKER.

620.113.2:658.562:31

To gain an insight into the various factors playing a part in the establishing of sampling inspection procedures some study of the mathematical foundations of the problem of sampling is needed. This leads to a systematic comparison of the different sampling-inspection systems available and to a choice amongst them.

The evaluation of the operating characteristic

In the previous article [1]) some fundamental aspects of sampling inspection have been considered and the three most common inspection procedures, single, double and sequential sampling, have been briefly discussed. It was found that the inspection performance of a sampling plan was duly expressed by means of the operating characteristic, a curve representing the probality of acceptance P of an inspection lot as a function of the percentage defective contained in it. The operating characteristics for three single sampling plans reproduced in *fig. 1* may serve to illustrate some of the fundamental properties of these curves.

As is to be expected, the operating characteristic shifts to higher percentages when the acceptance number c_0, that is the maximum number of rejects permitted in the sample, is increased. Furthermore the characteristic is steepest at the point where the percentage defective in the lot is about equal to the maximum permissible percentage of rejects, $c_0/n_0 \times 100\%$ (where n_0 is the sample size).

Lastly, if we keep the ratio c_0/n_0 constant while increasing the sample size, the curve remains at about the same place but becomes steeper.

In the limit, when the sample comprises the entire inspection lot, the operating characteristics will consist of a horizontal branch at $P = 1$ for percentages of defects less than the percentage of rejects permitted, and a second horizontal branch at $P = 0$ for percentages of defects greater than that value, these two branches being joined by a vertical. This is the operating characteristic corresponding to an ideal inspection in which all defects are detected and removed. In actual sampling, however, we inspect only a fraction of the lot and it is clearly of some interest to calculate the operating characteristic under such circumstances. To this end we shall have to resort to the calculus of probability.

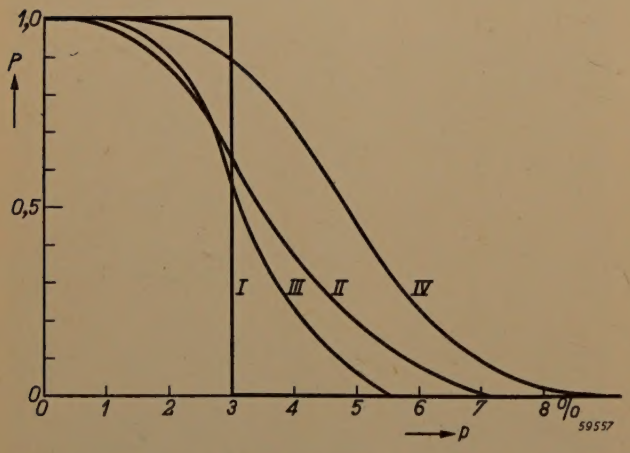


Fig. 1. Some operating characteristics for single sampling plans. *I* the characteristic when the sample size is equal to the size of the inspection lot; *II* the characteristic for a plan with a sample size $n_0 = 200$ and acceptance number $c_0 = 6$; *III* the same but for $n_0 = 400$, $c_0 = 12$. In these two cases the ratio c_0/n_0 has been altered but the sample size has been increased; as a consequence curve *III* is steeper than curve *II*. *IV* the characteristic for $n_0 = 200$, $c_0 = 9$. Compared with *II* the characteristic has shifted towards the right, owing to the increase in c_0, the sample size being the same in both cases.

Let us consider an inspection lot consisting of N pieces, M of which are defective; if from this lot we draw a random sample of size n_0, what will be the probability of finding m rejects in it? The correct formula expressing this probability is too cumbersome for numerical evaluation but fortunately an approximation can be used.

In most practical cases the sample is but a small fraction of the lot, so that we have

$$n_0 \ll N,$$

which automatically implies

$$m \ll M.$$

Moreover, in factory applications the percentage defective $p = M/N$ in the inspection lots is usually

[1]) H. C. Hamaker, Lot inspection by sampling, Philips Techn. Rev. **11**, 176-182, 1949 (No. 6), hereinafter referred to as I.

very small, of the order of a few percent; in other words we have

$$p \ll 1.$$

These conditions being satisfied, the probability of finding m defects in a sample of n_0 items is given by Poisson's formula:

$$\Pi(m; n_0, p) = \frac{e^{-n_0 p}(n_0 p)^m}{m!}. \quad . \quad . \quad . \quad . \quad (1)$$

According to this expression the probability depends on m and on the product $n_0 p$, but not on n_0 and p separately.

In single sampling, however, we are not so much interested in the probability of finding exactly m rejects in the sample as in that of finding c_0 or fewer rejects, which is obtained by summation as:

$$P(c_0; n_0 p) = \sum_{m=0}^{c_0} \frac{e^{-n_0 p}(n_0 p)^m}{m!}. \quad . \quad . \quad . \quad . \quad (2)$$

This is also the mathematical expression for the operating characteristic of a single sampling plan.

The probability of acceptance for double sampling, deducible in a similar way, leads to a somewhat more complicated formula, viz.:

$$P(c_1; c_2; c_3; n_1 p; n_2 p) = P(c_1; n_1 p) + \\ + \sum_{k=c_1+1}^{c_2} \Pi(k; n_1 p)\, P(c_3 - k; n_2 p). \quad . \quad . \quad . \quad (3)$$

Here the three criteria, c_1, c_2, and c_3, are so defined that an inspection lot is accepted when the number of rejects in the first sample is c_1 or less, and rejected when this number is greater than c_2, while acceptance after a second sample requires that the total number of rejects in both samples together shall be equal to or less than c_3 (see I page 179). On the right-hand side of (3) the first term measures the probability of accepting a lot after the first sample, and the second term measures the probability of accepting the lot after a second sample, no decision being reached in the first.

We are interested not only in the operating characteristic, but also in the average number of observations, or the average sample size, required per lot inspected. For single sampling this number is constant and equal to n_0, but for double sampling the situation is less simple. Sometimes one sample will suffice and sometimes two samples will be needed; and even if all the lots inspected contain exactly the same percentage defectives it still depends on the vagaries of chance whether one or two samples have to be drawn. All we can do, therefore, is to define an average sample size, and this is given by

$$\bar{n} = n_1 + n_2 \sum_{k=c_1+1}^{c_2} \Pi(k; n_1 p), \quad . \quad . \quad . \quad (4)$$

i.e. by the size of the first sample plus the product of the size of the second sample and the probability that a second sample has to be taken. As the formula shows, this average sample size depends on the percentage defective in the inspection lot. The fate of decidedly good or decidedly bad lots is usually determined by the first sample, and for such lots the average sample size is almost equal to n_1; but for lots of doubtful quality a second sample is often needed and the average sample size is consequently larger, so that as a function of the percentage defective it passes through a maximum.

The Poisson functions Π and P have been extensively tabulated [2]), and from these tables the operating characteristics for single sampling can at once be taken, while those for double sampling may easily be calculated.

For sequential sampling, too, formulae for the operating characteristic and for the average sample size have been derived, but as these are of a more complex character we shall not consider them in detail. They may be found in the writings of Barnard and of Wald [3]). Suffice it to say that, as with double sampling, the average sample size goes through a maximum as a function of the percentage of defects in the lot.

The fundamental parameters

In the concluding section of the previous article it was emphasized that the practical choice of a sampling plan is in the main influenced by economic factors which it is impossible to evaluate in a precise manner; and from this it was concluded that an accurate knowledge of the operating characteristic is not of essential importance, specification of the curve by a suitable set of parameters being all that is practically required. Now two important features of the operating characteristics strike the eye, namely the place where the curve shows its steepest descent, and the degree of its steepnes (see e.g. the characteristics of fig. 1). We shall use

---

[2]) E. C. Molina, Poisson's exponential binomial limit, D. van Nostrand, New York 1947. Less extensive tables in: E. L. Grant, Statistical quality control, McGraw Hill, New York 1946, and in T. C. Fry, Probability and its engineering uses, D. van Nostrand, New York 1928.

[3]) A. Wald, Ann. Math. Stat. **16**, 117-187, 1945 and J. Am. Stat. Ass. **40**, 277-306, 1945. G. A. Barnard, Supp. J. Roy. Stat. Soc. **8**, 1-27, 1946.

these two properties as a basis for the definition of two parameters which will play a fundamental part in our further investigation. To this end we shall locate the operating characteristic by its "point of control", p_0, defined as the percentage defectives corresponding to a probability of acceptance of $\frac{1}{2}$:

$$P(p_0) = \tfrac{1}{2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and we shall specify the steepness of the curve by its relative slope, h_0, in this point, defined by

$$h_0 = -\left\{\frac{p}{P}\frac{\mathrm{d}P}{\mathrm{d}p}\right\}_{p=p_0} = -\left\{\frac{\mathrm{d}\log P}{\mathrm{d}\log p}\right\}_{p=p_0}, \quad . \quad . \quad (6)$$

a negative sign being added to render h_0 an essentially positive quantity. Here the relative slope has been chosen instead of the absolute slope $(\mathrm{d}P/\mathrm{d}p)_{p=p_0}$ for a very special reason. As already pointed out, Poisson probabilities depend on the product np, and not on n and p separately. Consequently, by increasing or diminishing the sample size of a single sampling plan, or by altering the sample sizes of a double sampling plan both in the same ratio, while keeping the acceptance number c_0 or the criteria c_1, c_2 and c_3 at fixed values, the operating characteristic undergoes no other change than such as is equivalent to a change in the scale along the p-axis. Such a transformation does not alter the value of h_0, as is easy to see, so that this parameter is independent of the absolute size of the sample, or samples. This is a considerable advantage, as will appear farther on.

It also follows from the argument just given that any sampling plan can always be adjusted to some pre-assigned value of the point of control simply by changing the sample size or sizes. This is an adjustment that can always easily be effected, so that from now on we may consider all single sampling plans with the same acceptance number, or all double sampling plans with the same criteria and the same ratio of the sample sizes, as belonging to one class. Such a class may be represented by a single operating characteristic, which in its main features may be specified by one single parameter, namely its relative slope, h_0. All the operating characteristics of the sampling plans belonging to one class may be made to coincide in a simple and convenient way by plotting the probability of acceptance P not against the percentage defective p but against the ratio p/p_0.

Single sampling plans

For single sampling plans two simple empirical relations have been found to exist between the sample size n_0 and the acceptance number c_0 on the one hand, and the fundamental parameters p_0 and h_0 on the other, namely

$$n_0 p_0 = c_0 + 0{,}67\,, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and

$$\frac{\pi}{2} h_0^2 = n_0 p_0 + 0{,}06 = c_0 + 0{,}73 \quad . \quad . \quad . \quad (8)$$

The validity of the first of these equations was noted as early as 1923 by Campbell [4]), who calculated accurate values of $n_0 p_0$, some of which are reproduced in the second column of *table I*; comparison with the first column shows the high accuracy with which (7) is fulfilled.

For high values of c_0 equation (7) gives $n_0 p_0 \approx c_0$ or $p_0 \approx c_0/n_0$, a result that might have been expected: inspection lots with a percentage defectives equal to the percentage of rejects permitted in the sample should roughly have an equal chance of being accepted or rejected.

For large values of c_0, or of $n_0 p_0$, Campbell also deduced the approximate relation

$$\frac{\pi}{2} h_0^2 = n_0 p_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8a)$$

The empirical correction term $+0.06$ in (8) has the effect of rendering this equation valid down to the lowest values of c_0, as may be seen on comparing the last two columns in table I; the degree of approximation is highly satisfactory.

Table I. Values of $n_0 p_0$ and h_0 for single sampling plans with a given acceptance number.

c_0	$n_0 p_0$	h_0 from (8)	h_0 calculated from (2)
0	0.693	0.682	0.693
2	2.674	1.319	1.319
4	4.671	1.736	1.735
6	6.670	2.070	2.069
8	8.669	2.357	2.356
10	10.668	2.613	2.613
15	15.668	3.164	3.164
20	20.668	3.633	3.632
25	25.667	4.047	4.047
30	30.667	4.423	4.422

It goes without saying that with the aid of (7) and (8) the parameters p_0 and h_0 may be computed when n_0 and c_0 have been given, and vice versa.

Here, however, a slight difficulty arises owing to the fact that by their very nature n_0 and c_0 are positive integers, so that p_0 and h_0 cannot assume all the conceivable sets of values. This inconvenience may, however, be removed by the

[4]) G. A. Campbell, Bell System Technical Journal 2, 95-112, 1923.

artificial introduction of single sampling plans with broken values of c_0 and n_0. For example, a plan with $c_0 = 3.63$ is defined as follows. We set up a lottery between the numbers 3 and 4 such that the 4 has the probability of 0.63 of being drawn and the 3 the complementary probability of 0.37. Simultaneously with the taking of a sample we also draw a lot from this lottery, using the figure obtained as our acceptance number. Let P_3 and P_4 denote the probabilities of acceptance of an inspection lot when the acceptance numbers are 3 and 4 respectively, then the probability of acceptance for $c_0 = 3.63$ will be

$$P_{3.63} = 0.37\, P_3 + 0.63\, P_4.$$

Our definition is apparently equivalent to a linear interpolation between the operating characteristics corresponding to successive acceptance numbers.

Broken values of n_0 may be introduced in similar fashion. At first sight it may seem rather unnecessary and artificial to extend the definition of single sampling plans in the way we have just explained, and it is certainly not our opinion that plans with broken values of n_0 or c_0 should be of practical application. But the extended definition is of theoretical importance; since the operating characteristics for successive acceptance numbers form a regular sequence, we may conclude that equations (7) and (8), which were derived for integer values of n_0 and c_0, will hold good for broken values as well, so that by applying these equations we can from now on construct a single sampling plan corresponding to any set of positive values of the two parameters p_0 and h_0. This is of advantage when it is desired to compare different sampling systems, as we shall see in the next section.

Double sampling

Arguments set forth in the previous article suggested that a double sampling plan must be more efficient than a single sampling plan, in the sense that it may achieve the same inspection performance at the cost of a smaller number of observations. In what manner can we measure the gain thereby obtained?

To answer this question let us start by considering as a concrete example the double sampling plan with sample sizes $n_1 = 100$, $n_2 = 200$ and with the criteria $c_1 = 2$, $c_2 = 6$ and $c_3 = 10$. *Fig. 2* gives the operating characteristic of this plan as resulting from equation (3); precise calculation yields the following values for the fundamental parameters:

$$p_0 = 0.0379\,; \quad h_0 = 2.422.$$

Now it is natural to expect that the operating characteristics of two sampling plans having the same point of control and the same relative slope will be largely coincident, so that we are logically led to compare the double sampling plan under consideration with a single sampling plan with the same p_0 and h_0. According to (7) and (8) this requires

$$n_0 = 241.7 \quad \text{and} \quad c_0 = 8.49,$$

and this single plan corresponds to the charateristic indicated by the dots in fig. 2. The coincidence of the two characteristics appears to be almost complete, so much so in fact that for all practical purposes the inspection performances of the two plans may be considered as identical; we shall express this by calling them "equivalent".

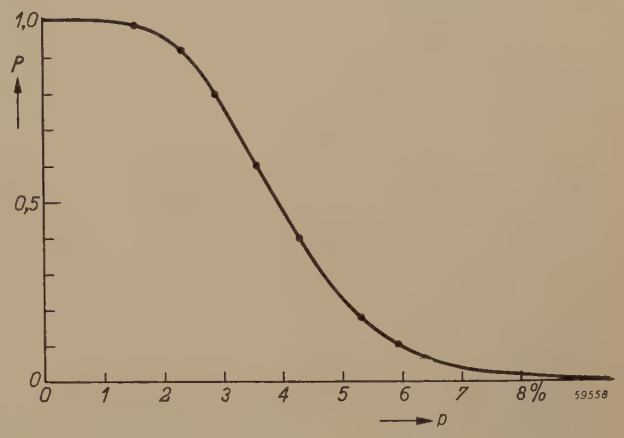


Fig. 2. Operating characteristic of a double sampling plan with $n_1 = 100$, $n_2 = 200$, $c_2 = 2$, $c_2 = 6$, $c_3 = 10$. The dots represent the characteristic of the equivalent single sampling plan, the curve that of the double plan.

Next, by calculating according to (4) the average sample size of the double sampling plan, we obtain curve *I* in *figure 3*, while the sample size of the "equivalent single sampling plan" is given by curve *II*. The advantage of double sampling is clearly demonstrated by these curves, but the choice of the sample sizes and hence of the scale on the left is still arbitrary. We shall arrive at a measure of the efficiency independent of this choice if we divide the average sample size of the double

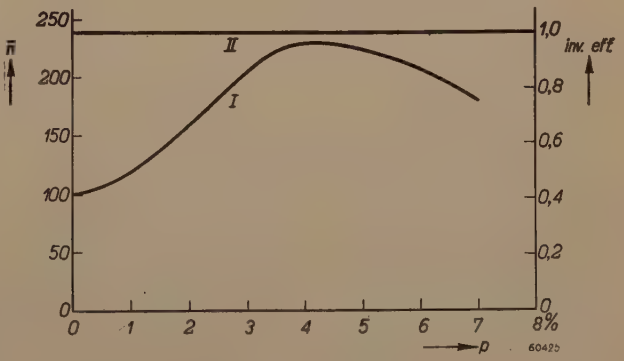


Fig. 3. The average sample size as a function of the percentage defective in the lot. Curve *I* gives the average sample size of the double sampling plan of fig. 2, which depends on the percentage defectives; curve *II* represents the (constant) sample size of the equivalent single sampling plan. For very good or very bad lots the average sample size of the double sampling plan is considerably smaller than the sample size of the single plan; the average amount saved is about 30%. Interpreted in terms of the scale on the right, curve *I* represents the inverse efficiency of the double sampling plan.

sampling plan by the sample size of the single sampling plan. The result is still represented by curve *I* in fig. 3, provided we use the scale on the right for its interpretation. The curve thus obtained measures the average number of observations required by the double sampling plan in terms of the number required by an equivalent single sampling plan, and may suitably be called the "inverse efficiency characteristic" [5]). We may read at a glance from this characteristic that for good batches the number of observations needed in double sampling is only about 40% of that needed in single sampling, whereas for lots of doubtful quality, near the top of the curve, the amount saved is relatively small. Further towards the right, beyond the range of the figure, the curve continues to descend until for very bad lots the inverse efficiency again reaches a value of 40%.

The reader will easily discern the general principle lying at the root of the example we have just discussed. By the extended definition of the previous section there exists one and only one single sampling plan for every set of positive values of the parameters p_0 and h_0. Since, moreover, of all sampling methods single sampling is by far the simplest, it is natural to adopt single sampling plans as a general standard of reference; the efficiency of any single sampling plan is then by definition equal to unity, while other sampling plans may be judged on the basis of their inverse efficiency, that is by comparing them with their single sampling equivalents. As a general principle this method of comparison will, of course, only be acceptable if the coincidence of operating characteristics is equally satisfactory as it was in fig. 2. The fact that this is generally so will become evident from the further cases we shall now proceed to consider.

As already pointed out, it is of some advantage to use a "class characteristic" obtained by plotting the ratio p/p_0 along the abscissa instead of p itself. The data needed for specifying such a class characteristic of a double sampling plan will then be four in number, viz. the ratio of the two sample sizes and the three criteria. Henceforth these will be written in the following notation:

$$D(n_2/n_1\,;\ c_1, c_2, c_3);$$

the symbol

$$D(2;\ 2, 6, 10)$$

for example signifying all double sampling plans for which the second sample is twice as large as the first and for which the criteria are 2, 6 and 10 respectively. The plan considered in figs 2 and 3 belongs to this class.

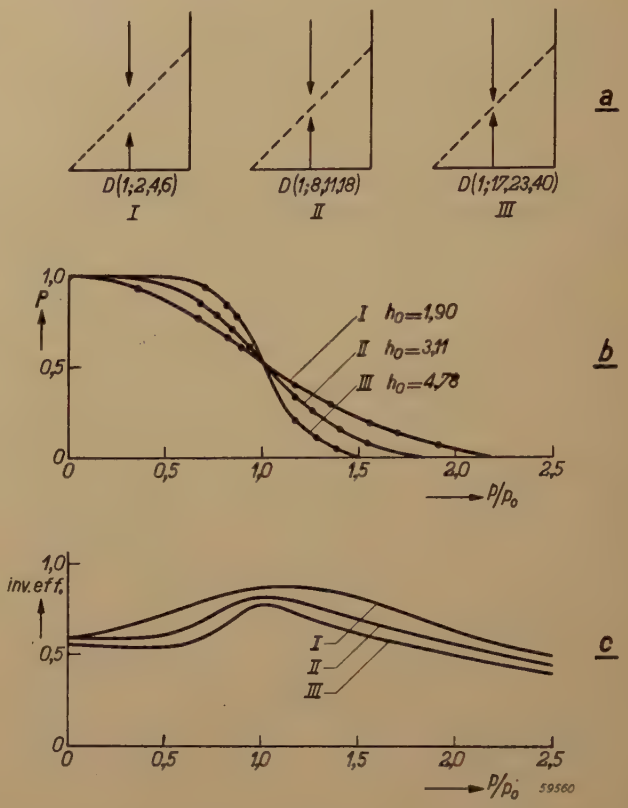


Fig. 4. *a*) Random walk diagrams corresponding to three double sampling plans with approximately the same configuration; *b*) The operating characteristics for these three plans, or the probability of acceptance as a function of the percentage defectives. The figure illustrates the high degree of concordance of the operating characteristics of double sampling plans with those of equivalent single sampling plans, the latter being indicated by the dots. *c*) Efficiency characteristics of the three plans under consideration, or the inverse efficiency as a function of the percentage defectives.

In *figs 4 a* to *c* three double sampling plans are compared for which the screens in the random walk diagram possess approximately the same geometrical configuration (see fig. 4*a*), but for which the criteria differ in magnitude. The coincidence of their operating characteristics with those of the equivalent single sampling plans is again highly satisfactory, as is illustrated in fig. 4*b*.

From the similarity of the configuration of the random walk diagrams we should intuitively expect the three plans to be approximately equal in efficiency, a conclusion corroborated by the efficiency characteristics in fig. 4*c*; only, the steeper the operating characteristic the sharper is the peak in the inverse efficiency curve. To remove this difference, which is more apparent than real, it may be of some practical advantage to plot the

[5]) The qualification "inverse" has been added because of the properties of these quantities; their inverse efficiency increases as the number of observations saved becomes less.

inverse efficiency not against p/p_0, as done in fig. 4*c*, but against the corresponding probability of acceptance P. The equivalent efficiency of the three plans under consideration is then brought to the fore much more conspiciously, as shown by

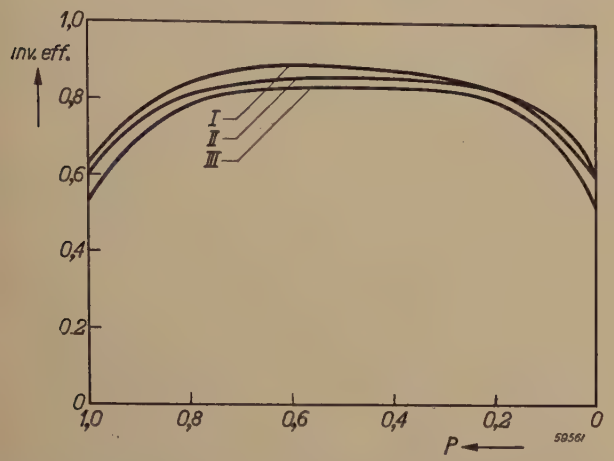


Fig. 5. Efficiency characteristics of the three plans of fig. 4; the inverse efficiency is now plotted as a function of the probability of acceptance P.

fig. 5. Moreover, in practice sampling plans are mostly used under such circumstances that about 95% of the lots submitted for inspection are accepted and only 5% rejected; in other words they are used in the neighbourhood of a probability of acceptance $P = 0.95$. A comparison of different sampling plans at this point can be most conveneintly performed with the aid of fig. 5. For example, we read at once from this figure that the saving in the number of observations attainable by double sampling is under practical conditions in the order of 25 to 30%.

According to equation (8) we have for single sampling approximately

$$p_0 n_0 \approx \sqrt{h_0}\,.$$

If, therefore, we wish to find a double sampling plan which, for a prescribed p_0, requires on the average the same number of observations as a given single sampling plan, the relative slope h_0 of this double plan may be chosen about $1/\sqrt{0.70} \approx 1.2$ times as high as for the single plan. This principle may sometimes be of service in changing over from one plan to another.

The most general double sampling plan requires five data for its specification, the two sample sizes and the three criteria, while in all existing sampling tables [6]) only four parameters are employed. This is because they use only those plans for which $c_2 = c_3$, a principle originally due to Dodge and Romig (see note [6])). From the random walk diagram in *fig. 6a* we may infer that this simplification implies that lots with fairly high percentages defectives will often be considered as doubtful after the first sample, while they will almost certainly be rejected after the second sample; hence the Dodge and Romig principle must be expected to lead to sampling plans with a relatively poor efficiency on the side of bad lots. This conclusion is affirmed by the efficiency characteristic *II* in fig. 6*b*. If, in addition to making $c_2 = c_3$, we increase c_1 so as to bring the opening in the first screen entirely above the line connecting the origin with the dividing point in the second screen (as in fig. 6*a* case *III*), this relative inefficiency is considerably enhanced. As shown by curve *III* in fig. 6*b*, the inverse efficiency of such a plan is greater than unity except for a small range of exceptionally good inspection lots, so that the number of observations needed is on the average greater than with an equivalent single sampling plan. Though there seems to be little sense in using plans of this kind they are nevertheless sometimes encountered in literature.

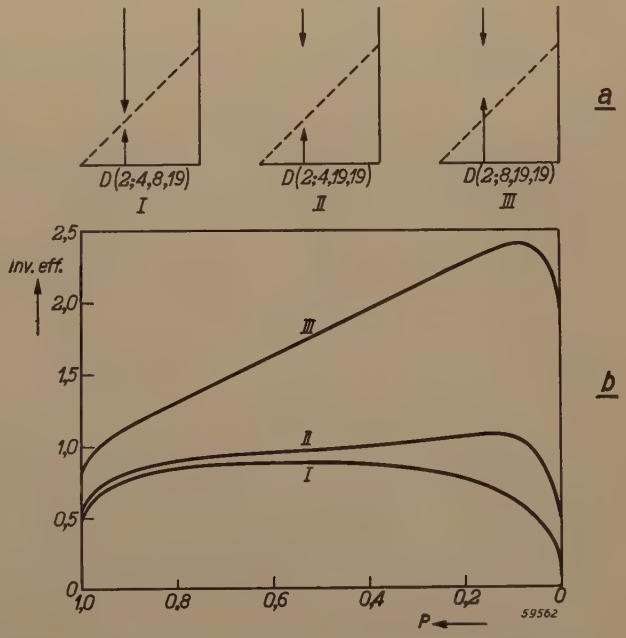


Fig. 6. Efficiency characteristics (*b*) of three double sampling plans corresponding to the three random walk diagrams (*a*). The pronounced influence of an incorrect choice of the criteria is clearly brought out, especially on the side of bad lots.

Case *I* in figs 6*a* and 6*b* on the other hand represents a plan where the Dodge and Romig principle has been abandoned and the criteria have been adjusted in the best possible manner.

[6]) H. F. Dodge and H. G. Romig, Single sampling and double sampling inspection tables, Bell System Technical Journal **20**, 1-61, 1941. Sampling inspection tables; single and double sampling, Wiley, New York 1944. Sampling inspection, Statistical Research Group Columbia University, Mc Graw Hill, New York 1948. See also E. L. Grant's book cited under reference [2]).

The total number of double sampling plans conceivable is of course unlimited, but from among them only a limited number are needed in practice, and it will by now have become clear that the efficiency characteristic may be helpful in guiding our choice and in weighing the advantages and disadvantages of different plans against one another.

A systematic study revealed that it is advantageous to make the second sample twice as large as the first, and that there is little point in altering this ratio. Furthermore, if for the sake of simplicity we adhere to the Dodge and Romig principle $c_2 = c_3$, it turns out that the achievement of a reasonable efficiency requires that c_3 must be 5 to 7 times as large as c_1. Also it was found convenient to classify double sampling plans according to their relative slope h_0; for the higher the value of h_0 the steeper is the operating characteristic and the larger the average sample size. On the basis of these principles we finally arrived at a set of double sampling plans as listed in *table II*.

Table II. Data for construction of a double sampling plan when p_0 and h_0 have been prescribed.

Plan	h_0	$n_1 p_0$
$D(2; 0, 1, 1)$	0.93	0.84
$D(2; 0, 2, 2)$	1.20	1.07
$D(2; 0, 3, 3)$	1.40	1.35
$D(2; 0, 4, 4)$	1.63	1.64
$D(2; 1, 5, 5)$	1.69	2.19
$D(2; 1, 6, 6)$	1.90	2.44
$D(2; 1, 7, 7)$	2.08	2.71
$D(2; 2, 10, 10)$	2.41	3.78
$D(2; 2, 12, 12)$	2.74	4.35
$D(2; 3, 15, 15)$	2.97	5.40
$D(2; 4, 20, 20)$	3.39	7.02
$D(2; 5, 25, 25)$	3.82	8.66
$D(2; 6, 30, 30)$	4.28	10.31

By way of illustration let us suppose that we wish to replace the single sampling plan $n_0 = 180$, $c_0 = 4$ by a double plan. From equations (7) and (8) we have $p_0 = 0.026$ and $h_0 = 1.74$ and from the second column in table II we see that the double plan D (2;1,5,5), with $h_0 = 1.69$, furnishes the best approximation. For this plan we have $n_1 p_0 = 2.19$, so that to fix p_0 at 0.026 we must take $n_1 = 2.19/0.026 = 84$, which in practice may be rounded off to 85. The double plan sought is now fully specified.

It would of course be possible to extend table II so that the successive values of h_0 lie closer together, but for practical purposes the set given is amply sufficient.

Sequential sampling

As explained in I, the random walk diagram of a sequential sampling plan is bounded by two parallel straight lines, one for acceptance and the other for rejection. Such a set of boundaries is completely specified by three parameters, for which we may choose, for instance, the angle α and the lengths of the two segments cut off from the ordinate axis (see *fig. 7*). Since we have found

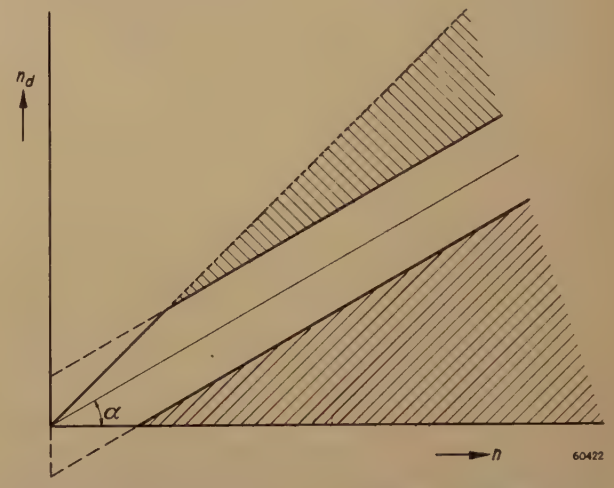


Fig. 7. Random walk diagram of sequential sampling plans. Only those sequential plans were found to be really efficient for which the line segments cut off from the vertical axis on both sides of the origin are approximately equal.

that in its main features an operating characteristic is fixed by two parameters only, p_0 and h_0, we are led to expect that for every set of values of p_0 and h_0 there exists a one-parametric set of corresponding sequential sampling plans. This is indeed the case.

By calculating the efficiency characteristics on the same basis as above it was found, however, that from such a set only those sequential plans are reasonably efficient for which the two decisive lines in the random walk diagram lie nearly symmetrical with respect to the origin. Furthermore, in the case of symmetry the equations specifying the boundary lines become extremely simple, namely

$$n_d = \pm h_0 + n p_0, \quad \ldots\ldots\ldots \quad (9)$$

where n_d denotes the number of rejects observed after inspecting n items. We go on inspecting as long as n_d remains within these limits.

From eq. (9) the construction of a sequential sampling plan with prescribed values of p_0 and h_0 becomes extremely simple. It must be added, however, that sequential sampling is frequently found to be impracticable, because one has to check for acceptance or rejection after every unit inspected (which might be remedied

by an automatic recording apparatus), while it may sometimes also happen that one has to inspect a very large number of items before a decision is reached.

Often it will be more convenient to steer a middle course by taking a limited sequence of concrete samples, and deciding between acceptance, rejection or a further sample only when the inspection of the sample at hand has been completed. Sampling plans of this type have been developed by the Statistical Research Group, Columbia University (see ref. [6])) and by J. H. Enters [7]). The method of application adopted by Enters is particularly simple, because he only varies the sample size, which in our terminology means that he uses only one value of h_0. It is possible to derive plans of the same simple form as the one he uses but with different values of the relative slope [8]), and these may be set out in just such a simple table as table II. However, our researches in this direction have not yet been completed.

Other sets of parameters

So far p_0 and h_0 have figured in our arguments as the two fundamental parameters, a practice that led to the simple equations (7) and (8) and to the successful introduction of the concept of inverse efficiency. But it does not necessarily follow that this set of parameters is also adapted to practical requirements. In all existing sampling tables which have been developed for practical application different parameters have been used. Some of these will now be considered in more detail.

The aim of sampling inspection is as a rule to protect the consumer against the delivery of bad lots occasionally produced. But at the same time it will be necessary to adjust the sampling plan in such a manner that the number of good lots perchance returned as unsatisfactory is not unduly large. We therefore have to take into account the requirements of both the consumer and the producer, and one set of parameters has been specially designed to achieve this purpose as directly as possible. Two points are chosen on the operating characteristics, one near the top and the other in the bottom part of the curve as shown in *fig. 8*.

The probabilities of acceptance for these two points are set at definite values, say at $P_1 = 0.90$ and $P_2 = 0.10$, and the corresponding percentages defectives, $p_{0.90}$ and $p_{0.10}$, are used as the two parameters by which the operating characteristic is fixed. The producer can then be guaranteed that of good lots with $p_{0.90}$ percent defectives not more than 10% are on the average rejected, while the consumer can rest assured that of lots with $p_{0.10}$ percent defectives no more than 10% are accepted. For this

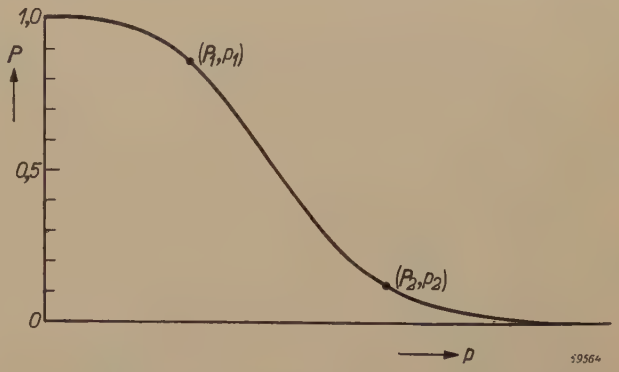


Fig. 8. Illustration of a set of parameters intended to take into account the requirements of producer and consumer. Two points on the operating characteristic are chosen for which the probabilities of acceptance have certain preassigned values, P_1 and P_2.

reason $p_{0.90}$ and $p_{0.10}$ have been called the 10 to 1 producers' and consumers' risk points; they have been employed by Wald and Barnard in their theoretical treatment of sequential sampling plans (see ref [3])), while Enters [9]) has constructed a nomogram from which the single sampling plan corresponding to given risk points can be read in a sample manner. This last nomogram may be further simplified and generalized. We then arrive at a simple graphical method by which the risk points corresponding to any set of values of p_0 and h_0 can be deduced in a very simple way, and vice versa. This method is described in the appendix to this paper.

In the "Army Service Forces" tables developed in the United States during the second world war, and with minor alterations subsequently taken over by the Statistical Research Group, Columbia University (see the tables mentioned under [6]) and Grant's book cited in reference [2])), the sample size has been adopted as one fundamental parameter together with one point on the operating characteristic, for which the 5% risk point of the producer was chosen ($p_{0.95}$). The reason for this was that in practice the sample size is often found to be the main limiting factor, so that it should more emphatically be taken into consideration. Again we may pass over from this set of parameters to p_0-h_0 and vice versa with the aid of a simple set of curves. For details we refer to the appendix.

[7]) J. H. Enters, T. Eff. Doc. **18**, 262-266, 1948 (No. 11).
[8]) J. H. Enters and H. C. Hamaker, not yet published.
[9]) J. H. Enters, The choice of the sample size in single sampling (in Dutch), Statistica **1**, 228-234, 1948.

Yet another parameter, already introduced in the earliest tables of Dodge and Romig and since then frequently employed, is the so-called "average outgoing quality limit", usually abbreviated as the AOQL. The argument leading to this concept runs as follows.

Suppose, firstly, that all inspection lots contain the same percentage defectives p and, secondly, that every lot rejected undergoes a 100% inspection and is then passed on to the consumer free of defectives.

If P denotes the probability of acceptance and p the percentage defectives in the lots then, under the said circumstances, a fraction P of the lots submitted for inspection is on the average accepted and a fraction $(1-P)$ rejected. Hence the average percentage defectives in the lots delivered is

$$p \cdot P + 0 \cdot (1 - P) = p \cdot P,$$

which is equal in value to the shaded area in *fig. 9*. We see at a glance from this figure that as p is made to increase, the average outgoing quality passes through a maximum, which is actually reached when the diagonal AC runs parallel to the tangent to the operating characteristic in point B. This maximum, called the "average outgoing quality limit", or AOQL, will be denoted by the symbol $\bar{p}_{max}$. The consumer can always be guaranteed that even under the most unfavourable circumstances, which never obtain in practice, the average percentage defectives in the lots he receives can never exceed the value of $\bar{p}_{max}$ corresponding to the sampling plan adopted.

As in the previous cases, a simple method for finding $\bar{p}_{max}$ for a given plan has been set out in the appendix.

The choice of a set of parameters

The principal sets of parameters to be found in literature have now been discussed. Which of these will prove most convenient in practice depends largely on personal taste and on the circumstances under which a sampling plan is operated. One aspect of the problem should, however, always be borne in mind, namely that by disposing of two parameters we have to serve three masters, the consumer, the producer and the inspector. It may be true that as a rule the inspector belongs to some department of either the producer or the consumer, but even so he will have only a limited personnel at his disposal, so that as far as the sample size goes he has an independent voice in the matter.

The producer is for obvious reasons always inclined to push the producers' risk point to the highest value possible, while the consumer is likewise inclined to maintain his risk point at the lowest level he can get accepted, so that if we adopt these risk points for our two parameters we are always likely to be driven to too steep an operating characteristic requiring an average sample size unacceptable to the inspector.

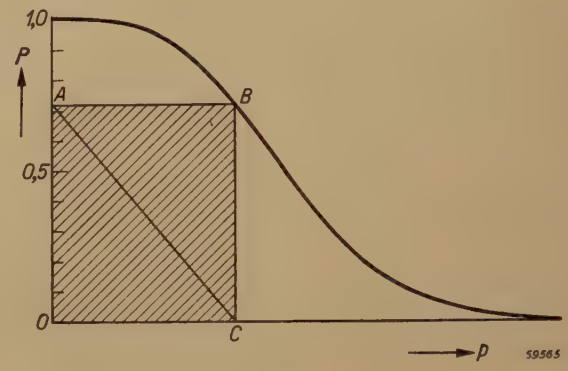


Fig. 9. Illustration of the concept of the "average quality limit".

If, on the other hand, we pitch, as in some of Dodge and Romig's tables, on the average outgoing quality limit and the consumer's risk point, we are using two parameters both intended to protect the consumer and we consequently run the risk of troubles with producer and/or inspector; or if, as in the tables of the Statistical Research Group, Columbia University, the leading part is allotted to sample size and producers' risk point, too little attention may be paid to the interests of the consumer.

Considered in this light the use of the point of control seems to offer some definite advantages. When producer and consumer have specified their respective risk points at say 2 and 5% the point of control should obviously be chosen about midway between them, say at 3.5%. By fixing p_0 in this fashion we are in a way simultaneously discounting the demands of both producer and consumer, while in the final adjustment of the slope h_0 we still possess a certain degree of freedom to satisfy the demands of the inspector. On these grounds the point of control p_0 and the sample size n_0 seem to constitute a particularly attractive set of parameters, the more so since the simple relation

$$n_0 p_0 = c_0 + 0.67$$

yields at once the acceptance number c_0 of the corresponding single sampling plan, while a double

sampling plan can easily be found with the aid of table II.

In a third article to appear in a subsequent issue of this journal we hope to explain how these principles have been employed in setting up a sampling table now in use on an extensive scale in the Dutch factories of our concern.

Appendix: The transition from one set of parameters to another

The transition from the 1-to-10 risk point to the fundamental parameters p_0 and h_0

As found above, sampling plans with the same relative slope h_0 possess operating characteristics that may be made to coincide by a simple adjustment of the sample size. It follows

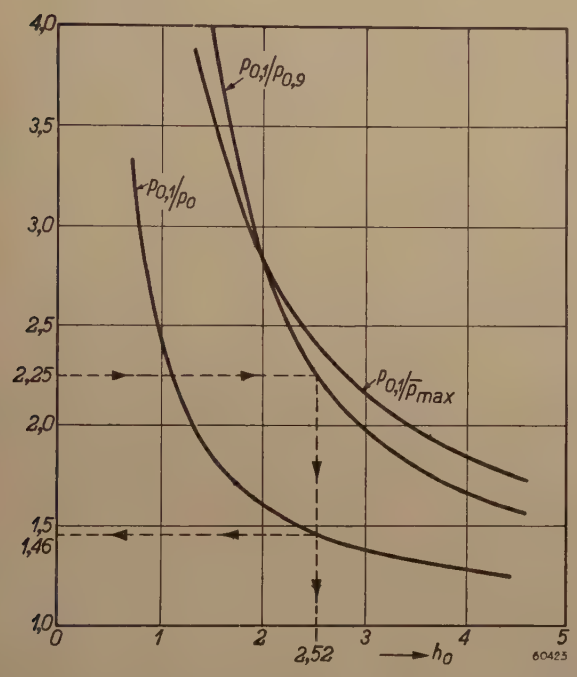


Fig. 10. Graph for the transition from one set of parameters to another. The broken lines with arrow points show how p_0 and h_0 can be found when $p_{0.10}$ and $p_{0.90}$ are known.

at once that the ratio $p_{0.10}/p_{0.90}$ is a function of h_0 alone, and the same holds true for the ratio $p_{0.10}/p_0$ [10]). By plotting these two ratios together as functions of h_0 we obtain a graph by means of which the transition from one set of parameters to the other can be readily achieved. If, for example, we have

$$p_{0.10} = 4.5\%, \quad p_{0.90} = 2\%,$$

and consequently

$$p_{0.10}/p_{0.90} = 2.25.$$

we start from the corresponding point on the ordinate in *fig. 10* and travel, as indicated by arrow points, first to the right until the curve $p_{0.10}/p_{0.90}$ is intersected and then downwards to the points of intersection with the $p_{0.10}/p_0$ curve and with the horizontal axis. On the horizontal axis we read

$$h_0 = 2.52,$$

and on the vertical axis

$$p_{0.10}/p_0 = 1.46,$$

whence

$$p_0 = 4.5/1.46 = 3.08\%.$$

Thereby p_0 and h_0 have been obtained. By substituting these values in (7) and (8) we find

$$c_0 = 9.25; \quad n_0 = 322,$$

which are rounded off to

$$c_0 = 9 \text{ and } n_0 = 320.$$

If a double sampling plan should be preferred table II provides D (2; 2, 10, 10) with $n_1 = 3.78/0.0308 \approx 123$ as the most suitable approximation.

The transition from n_0 and $p_{0.95}$ to p_0 and h_0

For single sampling the product n_0p_0, the ratio $p_{0.95}/p_0$ and consequently the product $n_0p_{0.95}$ depend on h_0 alone. A graph giving the two products n_0p_0 and $n_0p_{0.95}$ as a function of h_0 enables us to achieve our present purpose in a simple manner. Suppose we are looking for a sampling plan with

$$p_{0.95} = 2\% = 0.02 \quad \text{and} \quad n_0 = 130,$$

or

$$n_0p_{0.95} = 2.6.$$

Then, starting from the corresponding point on the vertical axis, we travel in just the same way through *fig. 11* as we did in fig. 10 (see the arrow points). We thereby obtain

$$h_0 = 1.90 \text{ and } n_0p_0 = 5.60; \text{ hence } p_0 = 5.60/130 = 0.043,$$
$$\text{and } c_0 = 5.60 - 0.67 = 4.93 \approx 5,$$

which fixes the single plan required. In deriving a double sampling plan on the same basis an approximation has to be accepted, because the average sample size now depends on the percentage defectives in the inspection lot and, consequently, on the circumstances under which the double sampling plan is operated. As mentioned earlier, however, double sampling roughly saves about 30% in the number of observations. A double plan requiring on the average about 130 observations per lot corresponds to a single sampling plan with a sample

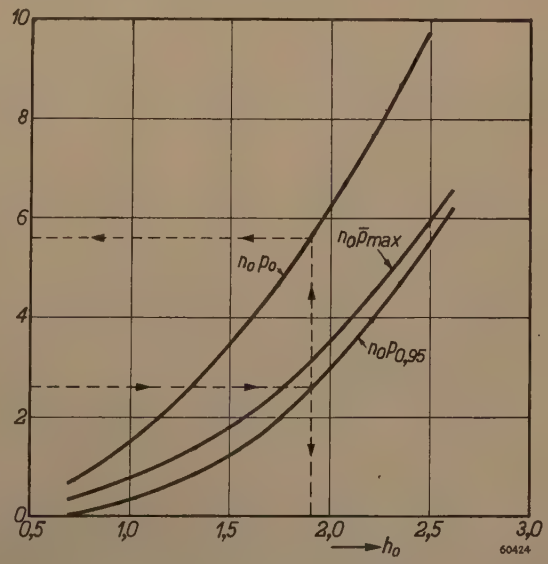


Fig. 11. As fig. 10. The arrow points indicate how h_0 and p_0 can be found for given n_0 and $p_{0.95}$.

[10]) In this connection it would be more consistent to write $p_{0.50}$ instead of p_0.

size of (100/70) × 130 = 185. Proceeding as before with $p_0 = 2\% = 0.02$ and $n_0 = 185$ we find $h_0 = 2.16$ and $p_0 = 4\%$, and table II indicates D (2; 1, 7, 7) with $n_1 = 2.71/0.04 = 68$ as the most suitable choice.

Determination of $\bar{p}_{max}$ for a given sampling plan

In analogy with the foregoing we may conclude at once that ratios such as $p_{0.10}/\bar{p}_{max}$ or $p_0/\bar{p}_{max}$ depend on h_0 alone, the same being true for the product $n_0\bar{p}_{max}$ in single sampling. This product and the first of the two ratios mentioned have also been plotted in figs 11 and 10 respectively, and, with the aid of the resultant curves, $\bar{p}_{max}$ can be found for a given sampling plan.

For example, if it is asked to find $\bar{p}_{max}$ for a single sampling plan with $n_0 = 130$, $c_0 = 8$, we have from (8) $h_0 = 2.36$ and then from fig. 11 $n_0\bar{p}_{max} = 5.23$, so that we must have $\bar{p}_{max} = 5.23/130 = 0.04$, which is the same value as that given in the tables of Dodge and Romig [6]). Proceeding in the same manner for the sampling plan $n_0 = 670$, $c_0 = 3$ we arrive at $\bar{p}_{max} = 0.0025$ as against 0.0029 according to Dodge and Romig, a difference of no practical consequence.

Conversely we read from Dodge and Romig's tables that the double sampling plan $n_1 = 150$, $n_2 = 350$, $c_1 = 1$, $c_2 = c_3 = 9$ corresponds to $\bar{p}_{max} = 0.011$ and $p_{0.10} = 0.03$. Consequently $p_{0.10}/\bar{p}_{max} = 2.7$ and with the aid of fig. 10 we find $h_0 = 2.08$, $p_{0.10}/p_0 = 1.58$, and finally $p_0 = 0.03/1.58 = 0.019$.

This provides the simplest method of finding p_0 and h_0 for the sampling plans given in Dodge and Romig's tables.

Summary. First, making use of Poisson's probability formulae, the expressions needed for a numerical theory of sampling plans are derived, and next two fundamental parameters are introduced for describing the main features of an operating characteristic, namely the "point of control" and the "relative slope". Between these two parameters on the one hand and the sample size and acceptance number for single sampling plans on the other, two simple relations, (7) and (8), are found to exist, as is illustrated in table I. The operating characteristics of double sampling plans and single sampling plans possessing the same values of the fundamental parameters are shown to coincide to a very high degree; such plans will give the same inspection performances in practice and may be called "equivalent". By comparing the average sample size of a double plan with the sample size of an equivalent single sampling plan one arrives at the concept of "inverse efficiency", which measures the relative number of observations that can be saved by resorting to the double sampling principle. Next the dependence of the inverse efficiency on the position of the screens in the random walk diagram is discussed and on the basis of these arguments double sampling plans are selected for practical use. These are set out in a table, from which a double plan with prescribed values of the fundamental parameters can easily be constructed.

A similar reasoning is then applied to sequential sampling and the resultant conclusions are briefly discussed.

Finally, various other parameters that have already made their appereance in literature are considered, such as the producers' and consumers' risk points, and the average outgoing quality limit. This leads to a discussion as to which of these parameters is to be considered most suitable in practice, and arguments are produced tending to recommend the point of control and the sample size as a particularly practical set.

In the appendix simple graphical methods are described which allow of a rapid transition from one set of parameters to another.

CONDUCTION PROCESSES IN THE OXIDE-COATED CATHODE

by R. LOOSJES and H. J. VINK.

537.311.33:621.315.592.4:
621.3.032.216

From the discovery of the thermionic emission of the alkaline earth oxides the modern oxide-coated cathode has been developed which is widely used in the field of electronics, for instance in radio valves, in cathode ray tubes and in gas-discharge lamps. An essential part of the emission process of this cathode is the electronic conduction of the oxide layer. Some new experiments and theoretical considerations concerning the nature of this conduction have made it possible to understand the behaviour of the oxide-coated cathode under various conditions better than has previously been the case.

Introduction

In 1903 Wehnelt discovered that a glowing strip of platinum covered with a small quantity of calcium oxide emits an appreciable quantity of electrons at temperatures where the thermionic emission of the metal itself is negligible. From this discovery in the course of time the modern oxide-coated cathode, as used, for instance, in most radio valves, has been developed, which consists of a porous sintered layer of mixed crystals of barium oxide and strontium oxide, 10-100 μ thick, on a metal core. The so-called indirectly heated cathode consists of a nickel tube coated on the outside with the oxide and surrounding a heating element (tungsten wire with an insulating coating). By means of this heating element the tube is heated to 1000-1100 °K and when the cathode is brought to a negative potential with respect to a collector electrode (anode) a strong thermionic emission takes place from the oxide coating.

At the temperatures mentioned a good oxide-coated cathode gives a thermionic emission (saturation current) of the order of 1 A/cm². The electrons emitted from the coating are restored from the metal core. From this it follows that the oxide coating itself must be conductive, and the object of the present article is to study this conduction process more closely, in the light of new experimental investigations.

Preparation of the oxide coating

As already remarked, with indirectly heated cathodes a tube of nickel is used as core for the oxide coating. For reasons which will be made clear later, the metal used is not pure nickel but an alloy obtained by adding small quantities of a readily oxidizable metal (Mg, Si, Mn, Al, Ti) to the nickel. First the metal core is coated with a thin layer of carbonates of alkaline earth metals (barium, strontium). These carbonates can be obtained, for example, by precipitation from an aqueous solution of the corresponding nitrates or hydroxides with ammonium carbonate. After being washed and dried the substance is ground for a number of hours in a ball mill with a solution of what is known as a binder (e.g. nitrocellulose dissolved in a volatile organic solvent). The resulting milkwhite suspension (size of particles 2-10 μ) is applied to the metal core, for instance, by means of a spray-gun. The solvent evaporates very quickly, leaving on the metal a porous coating of fine carbonate crystals cemented together and to the core by the binder.

The cathode is then mounted in the tube and heated in vacuo to 1100-1400 °K, whereupon first the binder decomposes and after that the carbonates decompose into oxides and carbon dioxide. The released carbon dioxide oxidizes the carbon of the binder to carbon monoxide. The gaseous products are pumped away, leaving a coating of pure oxides.

During the heating process the oxide coating sinters somewhat and adheres well enough to the core. The dissociation of the carbonate crystals requires great care. It must not be done too quickly nor at too high a temperature. If this is not properly attended to it may happen that the oxide coating sinters until it is no longer porous. Experience has taught that this is detrimental for the emission, for only a porous layer (porosity 50-60%) gives good emission. For the oxide coating mixed crystals of BaO and SrO (1 mol BaO to 1 mol SrO) are used because this material has proved to possess the best thermionic emission properties.

Activation of the cathode

The layer of earth alkaline oxides thus formed on a metal core is not capable of yielding the desired electron emission directly. To bring this about the cathode must first be "activated", and this is

generally done in the following way. The cathode is heated to 1100-1250 °K, while a voltage of some tens of volts positive with respect to the cathode is applied to the anode. Immediately after applying the voltage the emission is still very small, but it increases at first gradually and later on more rapidly. Ultimately the temperature of the cathode and the anode voltage have to be reduced in order to avoid very high emission currents, which appear to have a detrimental effect. After a few minutes a stationary state is reached and a cathode is obtained from which a continuous emission current of 10^3-10^4 A/m² (0.1-1 A/cm²) at a temperature of 1000-1100 °K and a pulse emission current (10^{-4} sec) of 2×10^4 to 10^5 A/m² can be drawn.

Although drawing current is mostly necessary for good activation, it is sometimes possible to activate the cathode without applying a voltage, i.e. without drawing current. This is particularly the case when the core contains the above-mentioned reducing materials (such as Si and Mg). Full activation can then be reached merely by heating the cathode in vacuo to a high temperature (1100-1200 °K).

This shows that a partial reduction of the oxide is essential for activating the cathode. It has been found that as a result of this reduction a certain quantity of free barium is formed in the coating of BaO-SrO mixed crystals. In the method of activation by drawing current described above this barium is released by electrolysis of the oxide coating, whereby Ba^{2+}-ions move in the direction of the metal core and O^{2-}-ions in the opposite direction. Finally the oxygen ions emerge from the layer in the form of free O_2 and the Ba^{2+}-ions are likewise neutralized, so that ultimately a quantity of free barium is left in the coating.

The quantities of barium in question are very small (in the order of 0.01%), but this free barium is nevertheless of essential importance for the emission. The fact is that as soon as the amount of barium is reduced, for instance by a momentary increase of temperature causing the barium to evaporate, or by chemical conversion (heating in a gaseous atmosphere containing traces of O_2, Cl_2 or of H_2O), the emission is greatly diminished. By introducing new free barium, for instance by evaporation, the emission is increased again.

The thermonic emission of a well activated oxide-coated cathode is, of course, strongly dependent on the temperature. In the case of the thermionic emission from metals (such as tungsten) Richardson's formula applies:

$$J_s = A\, T^2 \exp\,(-e\varphi/kT) \quad . \quad . \quad . \quad (1)$$

or [1]):

$$\log J_s - 2 \log T = \log A - 0.434\,(e\varphi/k)\,\frac{1}{T}, \quad (2)$$

where A represents a universal constant ($= 1.2 \times 10^6$ A/m² °K²) and φ is the work function amounting, for instance, to 4.5 V for tungsten ($e\varphi$ is the energy required to liberate an electron from the metal in vacuum). If, in analogy with this, one plots for an oxide-coated cathode $\log J_s - 2 \log T$ against $1/T$, a straight line

$$\log J_s - 2 \log T = \log A' - B/T. \quad . \quad . \quad (3)$$

is found. If B is then replaced by

$$0.434\, e\, \varphi'/k\,, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

we find the quantity φ', which in this case is also called the work function. For a well activated oxide-coated cathode φ' lies between 0.9 and 1.1 V.

When emission takes place electrons leave the oxide coating. In the stationary state this loss of electrons is compensated by electrons leaving the metal core and entering the oxide coating. Obviously a continuous thermionic emission is only possible if the electrons are able to pass through the coating into vacuum, i.e., if the oxide coating is an electronic conductor.

The oxide coating as semi-conductor

At first sight it may seem surprising that a layer of (Ba, Sr) O should be capable of conducting electrons, since it consists of transparent oxide crystals, and it is a known fact that such non-metallic substances generally behave as perfect insulators.

One would be inclined to ascribe this to electrolytic conduction of the oxides, and if the temperature is high enough the phenomenon of electrolysis indeed occurs; we have already seen from the foregoing that electrolysis plays a part in the process of activation. But the ionic conduction is far too small to account for the emission currents observed, and particularly so at the relatively low temperatures (1000-1100 °K) at which the cathode is used; it has been estimated that at such temperatures with a fully activated cathode the ionic current forms only 0.001% of the total current through the oxide coating, so that ionic conduction need not be considered.

It is known, however, that only absolutely pure crystals of stoichiometric composition are good insulators; impure crystals, contaminated in some

[1]) By "log" is to be understood the logarithm with the base 10.

way or other, for instance with foreign atoms or an excess of one of the constituent atoms, show conduction and are therefore called electronic semi-conductors.

The conductivity γ of an electronic semi-conductor depends strongly on temperature. A semi-conductor containing only one kind of impurity can be described by a formula of the type:

$$\gamma = K \exp(-E/kT), \quad . \quad . \quad . \quad (5)$$

where E represents a constant with the dimension of energy and K is likewise a constant which depends, however, on the amount of the impurity.

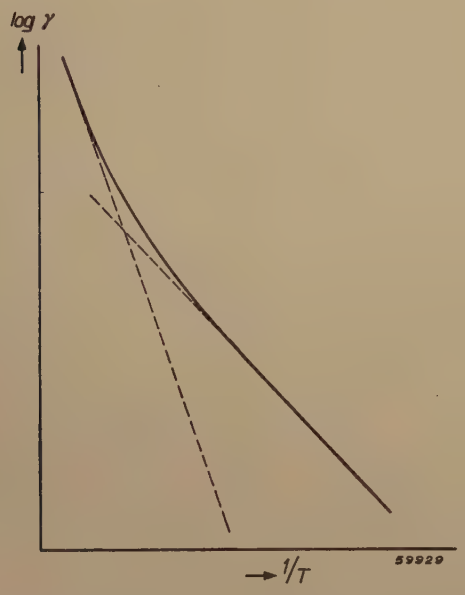


Fig. 1. The logarithm of the conductivity γ as a function of $1/T$ for a semi-conductor to which applies:

$$\gamma = K_1 \exp(-E_1/kT) + K_2 \exp(-E_2/kT)$$

(fully drawn line). The two dotted lines can be described by the separate terms.

When for such a material $\log \gamma$ is plotted as a function of $1/T$ a straight line is obtained from the slope of which the energy constant E can be deduced. It has been found experimentally that E is also slightly dependent on the impurity concentration, the value of E decreasing with increasing impurity concentration. Sometimes the semi-conductor contains more than one impurity, and when for instance there are two kinds of impurity γ can be described by a formula of the type:

$$\gamma = K_1 \exp(-E_1/kT) + K_2 \exp(-E_2/kT). \quad (6)$$

When in such a case $\log \gamma$ is again plotted against $1/T$ a curve is obtained as represented in *fig. 1*. From the slope of the two straight parts one can deduce E_1 and E_2.

Theory of electronic semi-conductors

Let us now consider the conduction and the thermionic emission of electronic semi-conductors more closely. According to the modern atomic theory (the quantum theory) electrons in the periodic field of the crystal lattice cannot have all possible energies. On the contrary, the possible energy levels are restricted to so-called allowed energy bands, whilst the intermediate values are prohibited [2]). In a crystal some of the allowed energy levels are always occupied by electrons while others are not. The lowest allowed bands are "filled", while the higher bands are "empty". Sometimes there is a half-filled band between the full and the empty ones, and this is the case with metals, the conductivity of which is due to this fact, since in this case the electrons in the half-filled band can be raised to a higher level in that band at the cost of very little energy taken from an external field, which means that an electric current flows through the metal.

If an entirely filled band is followed by an entirely empty band then the substance is an insulator, because the electrons cannot then be raised to a higher energy level, i.e. in the empty band, at the cost of little energy. This is only possible when there are partly filled levels between the filled and empty bands, as is the case with the above-mentioned crystals containing foreign atoms or, in the case of the oxide coating, with an excess of barium. These additional energy levels are usually at a small distance E_0 (in the order of 1 to 2 eV) below the empty band (*fig. 2*). Now it is possible that under the influence of the thermal agitation electrons are raised from the impurity levels to the empty band and become conducting electrons. Putting the number of impurity levels occupied by electrons (at $T = 0$ °K and per unit volume) as N, then according to the theory the number of electrons per unit volume (n) in the empty band at the temperature T is given by

$$n = 2 N^{\frac{1}{2}} \cdot \left(\frac{2\pi mkT}{h^2}\right)^{3/4} \exp\left(-\frac{E_0}{2kT}\right), \quad . \quad (7)$$

where m is the electron mass (9×10^{-31} kg) and h is Planck's constant (6.6×10^{-34} W sec^2).

For the conductivity we find:

$$\gamma = b n e, \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where b, the mobility of the conducting electrons, may be regarded as a temperature-independent

[2]) See for instance E. J. W. Verwey, Electronic conductivity of non-metallic materials, Philips Techn. Rev. **9**, 46-53, 1947

constant. When log γ is plotted as a function of $1/T$ in a certain, fairly narrow, temperature range then one obtains a straight line, since the term with $T^{3/4}$ has little effect compared with the exponential function. In this range γ can therefore be represented by the function

$$\gamma = K \exp(-E/kT), \quad \ldots \quad (9)$$

where E differs but little from $E_0/2$.

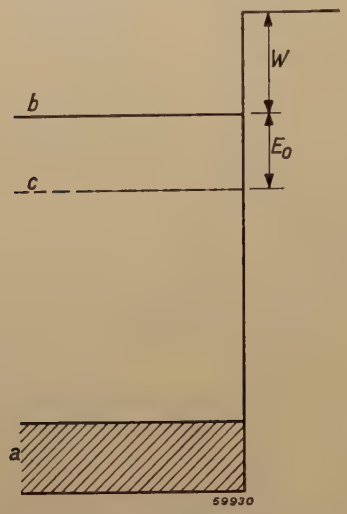


Fig. 2. Energy bands and additional levels of an electronic semi-conductor with an excess of metal atoms. *a* is the full band. E_0 is the distance between the narrow band of impurity levels (*c*) and the lowest level of the empty band (*b*). In the theory the factor $E_0/2$ plays an important part. W is the distance from the lowest level of the empty band to the zero level (vacuum).

A semi-conductor can also emit electrons. The foregoing theory is able to account for this and can describe the emission properties of the oxide-coated cathode. For a thermionic emission not only the energy separating the electrons in the impurity levels from the lower boundary of the empty energy band has to be overcome, but also the energy difference W between this lowest empty energy level and the "zero level" (corresponding to an electron with zero velocity in vacuo). For the thermionic emission of the semi-conductor one finds theoretically the formula

$$J_s = cne\left(\frac{kT}{m}\right)^{\frac{1}{2}} \exp(-W/kT), \quad \ldots \quad (10)$$

where c represents a numerical constant, viz. $(2/\pi^3)^{1/4}$. From the formulae (10) and (7) it follows that $e\varphi'$ (see (4)) differs little from $\frac{1}{2}E_0 + W$.

In what follows it will be investigated whether the conduction of the oxide coating can be explained from what is known about electronic semi-conductors. The results of the experimental investigation to be described will show that this is not fully the case.

Experimental investigation of conduction

Many investigations have been carried out regarding the conductance of the oxide coating, but hitherto it had not been possible to derive a coherent picture from the data obtained. What had been definitely established, however, was that almost exclusively electronic conduction takes place in the oxide-coated cathode.

For the investigations of this conduction carried out in the Philips Laboratories at Eindhoven a cathode was used as employed in cathode ray tubes. This cathode consists of a hollow cylinder of nickel (6 mm × 8 mm²) closed at one end and containing a heating element. The coating of Ba-Sr carbonate, 50 μ thick, covers the flat outer face of the closed end. Two of these cathodes, coated as smoothly as possible but otherwise quite normal, were pressed together with the carbonate coatings facing each other (*fig. 3*) and held in that position by springs. Thus between the two metal cylinders there was a carbonate coating 2 × 50 = 100 μ thick, which could be homogeneously heated by means of the two filaments. The temperature was measured with a chromel-alumel thermocouple.

The cathodes pressed together in this way were mounted in a glass bulb and this connected to a high-vacuum pump. Of course both the bulb and the cathodes were subjected to a thorough degassing process, so that ultimately, after dissociation of the carbonates and degassing of the whole unit, the bulb could be sealed with a vacuum of 10^{-5} mm Hg. In order to improve and maintain the vacuum, just before sealing the bulb a quantity of barium was evaporated as "getter", care being taken that this barium did not reach the cathode.

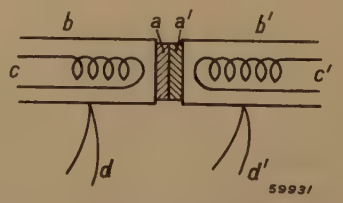


Fig. 3. Arrangement employed for measuring the conductance. *aa'* the two oxide coatings, *bb'* nickel tubes, *cc'* filaments, *dd'* thermocouples.

The core of the cathode consisted of ordinary "cathode nickel" containing Mg and Si, so that the oxide coating could be thermally activated. By carrying this out at relatively low temperatures it was possible to prolong the activation process over a number of hours, so that it could be interrupted at any moment to allow of measurements being taken in successive stages of the process. The conductance of the oxide coating approxi-

mately 100 μ thick was measured in various stages of activation as a function of temperature. These measurements were made with alternating current in a bridge circuit ("Philoscope" [3])), the voltage across the unknown resistance being about 1 V. At the same stages of activation current-voltage characteristics of the coating were recorded with the aid of short current pulses, the peak value of the voltage amounting to about 10 V.

Discussion of the results of the measurements

We shall first deal with the results obtained by measuring the conductance at low voltage: in *fig. 4* a graph is reproduced representing for a given oxide coating the logarithm of the resistance R ($-\log R$ is equal to $\log \gamma$ except for an additional constant) as a function of $1/T$ for various successive states of activation. It is clearly seen that as the activation increases, i.e. with increasing quantity of free barium in the coating, the conductance likewise increases and becomes less dependent on temperature. Further, in *fig. 5* another similar graph has been plotted, which shows that this curve can be divided into three distinct parts: part *I* (600-800 °K), only slightly dependent on temperature, part *II* (800-1000 °K), more strongly dependent on temperature, and part *III* ($>$ 1000 °K), again less dependent on temperature. Disregarding part *III*, one would be inclined to explain the conduction mechanism as being that of a semi-conductor in which two kinds of foreign atoms occur, for instance two kinds of barium atoms placed differently in the lattice, so that the conductivity could be represented by formula (6). The fact that the curves become less steep as the activation increases agrees with what has already been said about the effect of the number of impurity atoms on E. When the quantities E_1 and E_2 are calculated from the curves in accordance with formula (6) we find the following values for the successive stages of activation:

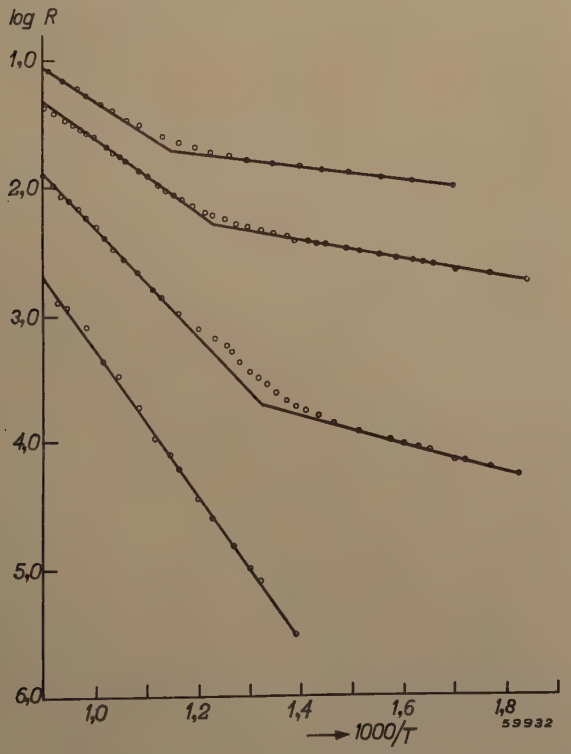


Fig. 4. The logarithm of the resistance R ($-\log R$ is equal to $\log \gamma$ except for an additional constant) as a function of $1/T$ for an oxide-coated cathode in successive stages of activation (from bottom to top).

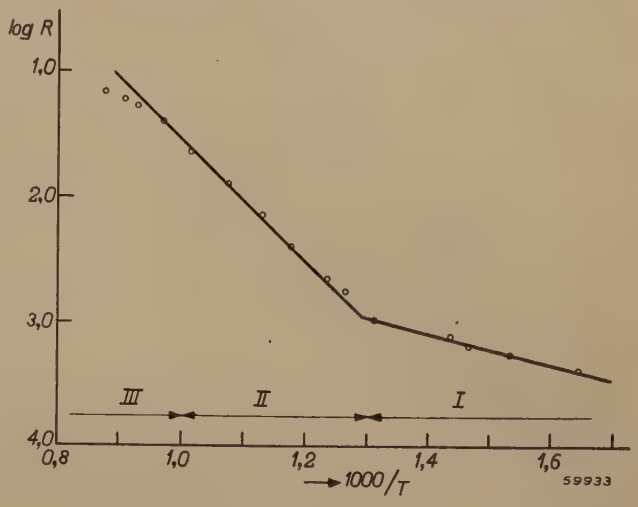


Fig. 5. The logarithm of the resistance R as a function of $1/T$ for an oxide-coated cathode. This curve clearly shows that there are three temperature ranges *I*, *II* and *III* in which the behaviour of the coating is different. The straight lines correpond to the dotted lines in fig. 1.

Stage	1	2	3	4	5	6
E_1	0.22	0.14	0.12	0.11	0.10	0.09 eV
E_2	0.98	1.14	1.24	0.96	0.96	0.94 eV
$e\varphi'$	1.07	—	1.10	0.90	1.02	0.87 eV

It appears that for all stages of activation E_2 is equal, within the limits of the accuracy of the measurements, to the product $e\varphi'$ of e and the work function at the respective stage. Now with indisputably real semi-conductors E has always been found to be much smaller than $e\varphi'$, so that the assumption of two kinds of impurity levels is improbable. Neither can the occurrence in part *III* of the log γ-$1/T$ characteristic be explained by the semi-conductor theory.

Something similar is encountered when studying the current-voltage characteristics, an example of which is given in *fig. 6*, from which it may be seen

[3]) See Philips Techn. Rev. 2, 270-275, 1937.

that with the J-V characteristics the same temperature ranges can be distinguished as those which played a part in the log γ-$1/T$ curves. Up to 700-800 °K the characteristics are straight, whilst between 800 and 1000 °K they curve towards the V-axis. Around 900 °K the curvature begins to

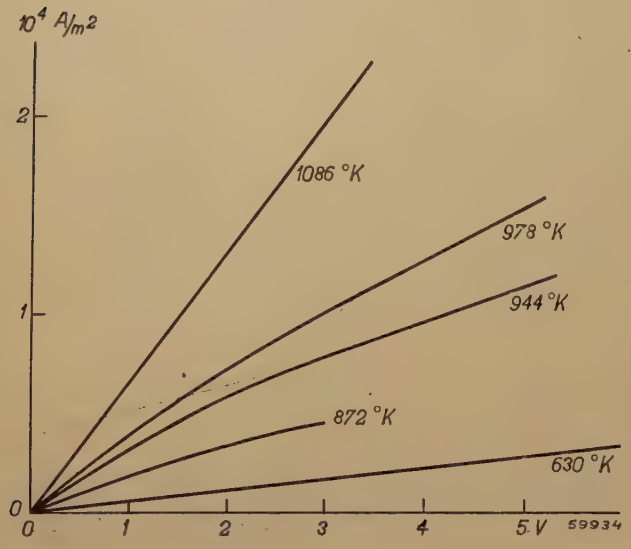


Fig. 6. J-V characteristics of an oxide-coated cathode at various temperatures. Between 800 and 1000 °K the characteristics are curved towards the V-axis.

diminish and above 1000 °K the characteristics are again straight. Now it is possible to explain a curved J-V characteristic by electronic conduction. For this it has to be assumed that rectifying contacts (barrier layers) exist at the boundary between the metal and the oxide coating. It is unlikely, however, that electrons of one kind of impurity levels (E_1) whose influence is preponderant at low temperatures ($<$ 800 °K) do not exhibit this effect, whilst those of another kind (E_2) whose influence predominates at higher temperatures ($>$ 800 °K) do show it. Whereas, therefore, there is nothing against ascribing that part of the conductance represented by the first term (the term with E_1) of formula (6) to real electronic conduction, for the behaviour represented by the second term (the term with E_2) some other explanation has to be found.

A new supposition regarding the causes of conductance

Attention has already been drawn in the foregoing to the fact that a good oxide-coated cathode has to be highly porous and that a closely sintered, non-porous cathode has poor emission. It is therefore surprising that hitherto so little attention has been paid to this question of porosity in trying to account for the phenomena in the oxide coating. It is our belief that the conduction by means of electrons present in the pores plays an important part in the phenomena observed. It is not illogical to suppose the existence of an electron gas of a certain density in the pores. For a long time it has in fact been accepted that owing to their emitting power the outermost grains of the oxide coating set up a cloud of electrons in a layer immediately adjacent to the surface of the cathode. If the outermost grains can do that, then why not the grains inside the coating too? The pores in the coating would then be filled with an electron gas coming from the electron clouds of the surrounding grains. As early as in 1918 von Laue [4]) investigated theoretically the electron density in a space bounded by walls emitting electrons. He proved that at low temperatures the density varied with temperature in the same way as the emission. At higher temperatures, and thus with increasing emission, owing to the mutual repulsion of the electrons, the density of the electron gas in the centre of the cavity diminishes as compared with the density close to the emitting surface. As a matter of fact the density in the centre of the cavity increases only in direct proportion to the temperature, whilst that close to the wall always increases according to the same exponential law as the emission itself. We shall revert to this presently, but here we can already find support in von Laue's results for our conceptions regarding the part played by the pores in the conduction process.

The pores referred to above should not be imagined as being cavities separated from each other, because then the electrons would always have to pass through a semi-conducting layer to move from one cavity to the next. The cavities are to be regarded more as a coherent sequence, so that winding channels are formed in the material, leading from the metal core to the free cathode surface. We shall now see what effect this has on the conduction.

In these channels the electrons move with a mean free path l, which is of the order of magnitude of the cross section of the channel, i.e. of the diameter of the grain. When a field F is applied, during the time t taken by the electrons to travel their mean free path l these electrons (with charge e and mass m) undergo an acceleration

$$a = \frac{eF}{m}. \quad \ldots\ldots\ldots \quad (11)$$

[4]) M. von Laue, Jahrb. Radioaktivität und Elektronik **15**, 205-256, 1918; Ann. Physik **58**, 695-711, 1919.

During this time the average increase in velocity in the direction of the field is

$$\overline{\Delta v} = \frac{eF}{2m} t .$$

We now distinguish between two cases:
1) The average velocity $\bar{v}$ is great compared with $\overline{\Delta v}$. In this case $t = l/\bar{v}$ and therefore

$$\overline{\Delta v} = \frac{e}{2} \frac{F}{m} \cdot \frac{l}{\bar{v}} \quad \ldots\ldots\ldots \quad (12)$$

Since the acceleration is destroyed after the electrons have traversed the mean free path, $\overline{\Delta v}$ represents at the same time the average velocity of the electron in the direction of the field, so that for the current density we find:

$$J = n e \overline{\Delta v} = n \frac{e^2 F}{2m} \cdot \frac{l}{\bar{v}} , \quad \ldots \quad (13)$$

where n represents the number of electrons per unit volume. A similar formula was derived long ago by Drude to explain the electronic conduction of metals. Since J is proportional to the field strength F, and this in turn is proportional to the applied voltage V, this formula yields a rectilinear J-V characteristic (Ohm's law).
2) The average velocity $\bar{v}$ is small with respect to $\overline{\Delta v}$. This is the case if l or F is sufficiently large. Then one can write for t: $t = l/\overline{\Delta v}$ and thus

$$\overline{\Delta v} = \sqrt{\frac{eF}{2m} l}$$

and

$$J = ne \sqrt{\frac{eF}{2m} l} \quad \ldots\ldots \quad (14)$$

Since F is proportional to the potential difference applied, Ohm's law no longer holds in this case and we get a parabolic J-V characteristic, curved towards the V-axis.

The possiblility of this being the case with the oxide-coated cathode may be seen from the following consideration. With a potential difference of 10 V across a coating 100 μ thick the increase of the kinetic energy per mean free path (of the order of magnitude of the grain, thus 2-10 μ) is 0.2-1 eV. This is very much greater than the average thermal kinetic energy ($^3/_2\, kT$), which at 800 °K amounts to 0.1 eV. The fact that at very high temperatures ($>$ 1000 °K) the characteristic again straightens out can be explained by the theory of von Laue, according to which the electrons are driven out of the centre of the cavities (*fig.* 7) owing to the space charge. As a result the electrons move mainly in a thin layer along the wall of the pores, because an electron leaving the wall does not travel across the cavity but is reflected by the space charge back to the wall on the same side. Thus the free

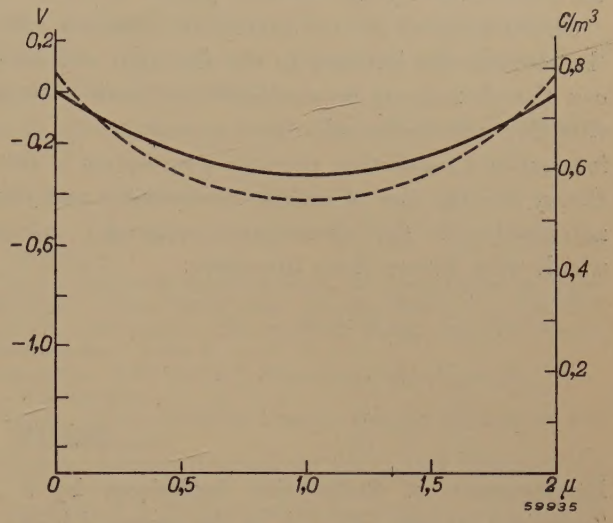


Fig. 7. Potential V in volts (fully drawn line) and charge density in coulombs/m³ (dotted line) between two thermionic plates at a distance of 2μ (T = 952 °K, $J_s = 3.8 \times 10^4$ A/m²).

path is greatly reduced and we then again have the case of formula (11), the characteristic becoming straight again. Since, moreover, the average density in the layer referred to is less than the boundary density, which is proportional to the saturation current, it also follows that with increasing temperature the conduction in the pores within this temperature range does not increase so much as it does in the temperature range where the space charge is still negligible and where, therefore, the conduction increases with the emission.

In order to ascertain whether the explanation given for the curved J-V characteristic is correct we calculated J as a function of V according to formula (14) for a cathode which at 1000 °K should have an emission of 8×10^4 A/m² with $\varphi' = 1.1$ V. The electron density is taken for a simple case (two infinite, parallel, flat plates at a short distance from each other). For a distance of 2 μ we found $J = 9.0 \times 10^3 \sqrt{V}$ A/m² and for a distance of 5 μ $J = 8.5 \times 10^3 \sqrt{V}$ A/m². These current densities are of the same order as those found experimentally.

Summarizing it may therefore be said that in the porous oxide coating the conduction by the electron gas between the grains is of essential importance. Without taking this conduction into account it is not possible to give a satisfactory explanation of all the phenomena found experimentally. The conduction by the electron gas and the electronic conduction of the crystals of the oxide coating are

to be regarded as two parallel processes which together determine the conduction of the coating. At low temperatures conduction through the crystals predominates and at high temperatures the conduction by the electron gas.

Strong support for the theory of conduction by the electron gas is found in the fact that not only can the phenomena be predicted qualitatively but that they can also be calculated quantitatively. It is interesting to note that the only parameters in this theory are the size of grains, the porosity and the magnitude of the thermionic emission, values which were known from literature.

Summary. A description is given of the methods of preparing and activating an oxide-coated cathode, from which it appears that the carbonate coating originally applied has to be subjected to such a heating process as to result in a loosely sintered, porous oxide coating. The electrons emitted from the coating have to be restored by conduction through the oxide coating. It is shown to be probable that the conduction of the oxide coating is due to two mechanisms acting in parallel: the electronic conduction of the grains, predominating at low temperatures (< 800 °K), and the conduction through the electron gas in the pores between the grains, which is preponderant at high temperatures (> 800 °K). With the aid of the conduction through the pores it can also be seen why the J-V characteristic for high voltages curves towards the V-axis. The fact that at the highest temperatures (> 1000 °K) the conduction lags behind the emission, as also the fact that in this temperature range the curvature of the J-V characteristic disappears, can be explained by taking into account the effect of the space charge upon the electron density and upon the field in the pores.

BOOK REVIEW

Fundamentals of Radio-valve Technique, by J. Deketh, 535 pages, 384 illustrations. — Philips' Technical Library (Book I of the series of books on electronic valves) — Published by N.V. Philips' Gloeilampenfabrieken, Technical and Scientific Literature Department, Eindhoven, Netherlands, 1949.

In every field of science or technical engineering there is a need not only of a detailed handbook for the specialist but also of a more concise source of information serving as a general guide for the student and at the same time as a book of reference for others indirectly concerned with the subject matter. Without being written in such a popular language as to make it unnecessarily superficial, such a book must nevertheless be comprehensible and, moreover, bring forward all the essential points in a logical manner.

For the author of the book discussed here on the fundamentals of radio valve technique this difficult task was perhaps made somewhat easier by reason of his working in surroundings where he was continually in contact with various classes of people engaged in this particular field of practical science. In those very same surroundings it appeared that a book such as this is appreciated not only by young technicians but also by those who are mainly interested in chemistry and physics and who need to know something about electronic systems in connection with their own particular work.

This book deals exclusively with radio receiving valves and their application in radio receivers, thus not with transmitting valves, cathode ray tubes, and so on.

The first three chapters are devoted to the fundamental principles upon which the working of electronic valves is based. Chapters IV and V go deeper into the matter of the thermionic emission. The next three chapters are concerned with the technology of the valves, a subject which has been gone into at some length, and rightly so, because also in the application of radio valves this technology influences the possibilities and limitations.

Following upon a review of the various functions of the electronic valve in a radio receiver and of the types of valves that have been developed for reception purposes, in chapters XII to XVI the author deals with the characteristic properties of these valves, whilst in the next seven chapters it is explained how these properties can be utilized in the various functions, such as low-frequency amplification, high- and intermediate-frequency amplification, detection, etc. Thus a complete insight is given into the working of a radio receiver, without, however, dealing in detail with other components like capacitors, inductors, loudspeakers, etc.; this limitation has undoubtedly contributed towards the sound treatment of the subject matter.

The advantage that the author had from his close contact with a large factory making radio valves and receivers is manifest from the space that has been reserved in this book to the treatment of phenomena of interferences such as noise, hum, microphony, secondary emission and the ageing of valves. These are points which certainly have to be taken into account in the designing and developing of receivers, and the reader will find, in a somewhat condensed form, a great deal of information about these phenomena which will enable him to evaluate for himself the extent of all sorts of effects.

An appendix of about 50 pages gives definitions, formulae and tables of great value to the technician in his daily work.

Finally, an extensive bibliography is given at the end of the book, whilst also in the text, where necessary, references are made to books dealing with a particular detail at greater length.

Naturally there is not much scope within the confines of such a book to give the full derivation of all sorts of formulae. As a rule indispensable formulae, like those for emission, anode current, noise voltage, etc., are given direct. Where formulae have been obtained by derivation the reader is assumed to have a knowledge of algebra, goniometry and the principles of differential calculation. Further, of course, a reasonable electrotechnical grounding is required.

H. van Suchtelen.

The books of Philips Technical Library are distributed in:

Norway, Sweden, Holland, Indonesia, Dutch West Indies: Meulenhoff & Co N.V., Beulingstraat 2-4, Amsterdam
Denmark: Jul. Gjellerups Boghandel, Sølvgade 87, København K.
Finland: Akateeminen Kirjakauppa, 2 Keskuskatu, Helsinki
Germany (Trizone): Buch- und Zeitschriften Union, Harvestehuder Weg 5, Hamburg 13
Switzerland: A. Francke A.G., Bubenbergplatz 6, Bern; S. A. Payot, Lucerne
Belgium: N.V. Alg. en Techn. Boekhandel, v.h. P. H. Brans, Prins Leopoldstr. 28 - Borgerhout, Antwerp; N.V. Standaard-Boekhandel, Huidevetterstraat 57-59, Antwerp
Luxemburg: Librairie Paul Bruck, 50 Grande Rue, Luxemburg
France: Maison Dunod, 42 Rue Bonaparte, Paris VI
Great Britain and Eire: Cleaver Hume Press Ltd., 42 A South Audley street, London W. 1.
Spain: Editorial Pueyo, Tetuan 5, Madrid
Portugal: Livraria Bertrand, Rua Garrett 73, Lisbon
Italy: Librairie Internationale Corticelli Via S Tecla 5, Milann
Austria: Buchhandlung Minerva, Mölkerbastei 4, Vienna I
Yugoslavia; Jugoslovenska Knjiga, Marsala Tita 32, Belgrade
Greece: "Eleftheroudakis", Constitution Square, Athens
Turkey: Librairie Hachette, 469 Istikalât, Caddesi, Beyoğlu, Istanbul
Egypt: Lehnert & Landrock Succ., 44 Sherif Pasha str., Cairo
Palestine: Pales Press Ltd., Allenby Road 119, Tel Aviv
U.S.A. and Canada: Elsevier Book Comp. Inc., 215 Fourth Ave., New York 3
Argentine: Editorial Sud Americana S.A., Alsina 500, Buenos Aires
Brasil: Livraria Editoria Kosmos, Rua do Rosario 135-137, Rio de Janeiro
Uruguay: Libreria International S.R.L., Calle Uruguay 1331, Montevideo
Venezuela: C.A. Philips Venezolana, Apartado 1167, Caracas
Australia: Philips Electrical Industries of Australia (Pty), Ltd., "Philips House", 69-73 Clarence Street, Sydney
New-Zealand: Philips Electrical Industries of New-Zealand Ltd., G.P.O. Box 1673, Wellington G. 1.
South-Africa: Technical Books & Careers, 37 High Court Buildings, 17 Joubert Street, Johannesburg
India: Philips Electrical Co., (India) Ltd., "Philips House", 2 Heysham Road, Calcutta 20
Pakistan: Philips Company of Pakistan Ltd., Bunder Road, P.O. Box 301, Karachi 3
Syria and Lebanon: Philips Liban-Syrie S.A., P.O.B. 670, Beyrouth

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the address printed on the back cover.

1862: P. C. van der Willigen: Contact arc welding (Sheet Metal Industries **26**, 393-398, 402, 1949, Febr. No. 262).

A survey is given of the new method of arc-welding with contact electrodes. Constant arc length, independency of physical conditions of the welder, ease of starting of the arc, ease of welding and more reliable welds are the chief advantages. Other features are a special form of penetration and greater welding speed. Special application such as contact-arc spotwelding and under-water welding are briefly dealt with. See Philips Techn. Rev. **8**, 161-166 and 304-308, 1946.

1863: J. L. Snoek: The Weiss-Heisenberg theory of ferromagnetism and a new rule concerning magnetostriction and magnetoresistance (Nature London **163**, 837, 1949, May 28).

The Heisenberg model of ferromagnetism, in contradiction to the band theory, leads one to expect that the exact filling up of lattice points with an equal number of electrons will make itself felt in some properties of a ferromagnetic alloy. From available data and from special experiments it is shown that under these conditions the magnetostriction is zero and the change in the electric resistance due to the applications of a magnetic field passes through a maximum.

1864: J. H. van Santen and F. de Boer: High indices of refraction of barium titanate and other heteropolar compounds (Nature London **163**, 957, 1949, June 18).

It is shown that the polarizability of the oxygen ions in $BaTiO_3$ and similar compounds as derived from a corrected Lorentz-Lorenz formula (see these abstracts, No. **1833**) is mainly due to a low value of the characteristic frequency ν_e occurring in the expression $c/(\nu_e^2-\nu^2)$.

An estimation of ν_e on the basis of lattice data is given. At the same time a qualitative explanation is given of the high refractive index of compounds containing titanium ions surrounded by six oxygen ions.

1865: H. Rinia: The Schmidt optical system (Bull. Schweiz, Elektrotechn. Verein **40**, 580-585, 1949 No. 17).

In this article the principle of the Schmidt optical system is outlined. It is shown that the conventional form gives fifth order coma for low magnifications. Means are indicated to compensate this coma. The cause and magnitude of the lateral spherical aberration are discussed and a new method to compensate this aberration is shown.

1866*: J. D. Fast: The influence of oxygen, nitrogen and carbon on the impact strength of iron and steel (reprint No. 7, Int. Foundry Congress 1949, Amsterdam).

Description of method of preparation of reproducible samples of pure metals charged with known quantities of oxygen, nitrogen and carbon.

1867: G. H. Jonker and J. H. van Santen: Properties of barium titanate in connection with its crystal structure (Science **109**, 632-635, 1949, June 24).

The behaviour of the permittivity of barium titanate and related compounds as a function of temperature is extensively discussed. (See these abstracts, No. **R92** and Philips tech. Rev. **11**, 176-186, 1949, No. 6.)

1868: W. Hoogenstraaten and F. A. Kröger: The intensity-dependence of the efficiency of fluorescence of willemite phosphors (Physica **15**, 541-556, 1949, No. 5/6).

The efficiency of the fluorescence of zinc silicate activated with a high concentration of manganese is dependent on the temperature and on the exciting intensity when excited by short-wave ultra violet. The efficiency is constant over a wide range of low exciting intensities, increases at intermediate exciting intensities and will probably tend to a constant value again at high exciting intensities. Similar effects are observed with Zn_2SiO_4-Mn_2SiO_4 in which small quantities of Fe_2SiO_4 have been incorporated. This behaviour can be explained on the basis of the theory of hole-migration for bimolecular phosphors of Schön-Klasens, if the roles of positive holes and excited electrons in that theory are interchanged. Values of the activation energies for migration from fluorescence centres and killer centres are computed, and the theoretical temperature-dependence curves calculated from these data for different intensities are shown to be in accordance with the observed curves.

1869: F. A. Kröger and W. Hoogenstraaten: Temperature quenching and decay of fluorescence in zinc-beryllium silicate activated with manganese (Physica **15**, 557-588, 1949, No. 5/6).

Zinc silicate and zinc-beryllium silicate activated by manganese contain various types of fluorescence centres which differ in the wavelength of the emitted radiation, the probability of the fluorescence transition, and the vibrational interaction with the surroundings, which regulates the dissipation of energy of excitation. The centres are assumed to be manganese ions with different numbers of the other manganese ions at neighbouring sites (clusters). Quenching of fluorescence is due to radiation-less processes starting from the excited state of the centres (characteristic quenching). If particular conditions are fulfilled, however, quenching may also occur from (higher) excited states of a different type. In this case energy of excitation is transferred to centres especially suited for the dissipation of energy (quencher centres). This type of quenching is found in zinc silicates at medium manganese concentrations, and in zinc- and zinc-beryllium silicates containing foreign quenchers (iron). It is assumed that quencher centres are fluorescence centres with a low quenching point.